A STUDY OF QUANTUM ANNEALING DEVICES FROM A CLASSICAL PERSPECTIVE

A Dissertation

by

ANDREW JOSEPH OCHOA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Helmut G. Katzgraber
Committee Members,	Stephen A. Fulling
	Joseph H. Ross
	Kim-Vy H. Tran
Head of Department,	Peter M. McIntyre

December 2017

Major Subject: Physics

ABSTRACT

Spin glasses are experiencing a revival due to applications in quantum information theory. In particular, they are the archetypal native benchmark problem for quantum annealing machines. Furthermore, they find applications in fields as diverse as satisfiability, neural networks, and general combinatorial optimization problems. As such, developing and improving algorithms and methods to study these computationally complex systems is of paramount importance to many disciplines. This body of work attempts to attack the problem of solving combinatorial optimization problems by simulating spin glasses from three sides: classical algorithm development, suggestions for quantum annealing device design, and improving measurements in realistic physical systems with inherent noise. I begin with the introduction of a cluster algorithm based on Houdayer's cluster algorithm for two-dimensional Ising spin-glasses that is applicable to any space dimension and speeds up thermalization by several orders of magnitude at low temperatures where previous algorithms have difficulty. I show improvement for the D-Wave chimera topology and the three-dimensional cubic lattice that increases with the size of the problem. One consequence of adding cluster moves is that for problems with degenerate solutions, ground-state sampling is improved. I demonstrate an ergodic algorithm to sample ground states through the use of simple Monte Carlo with parallel tempering and cluster moves. In addition, I present a non-ergodic algorithm to generate new solutions from a bank of known solutions. I compare these results against results from quantum annealing utilizing the D-Wave Inc. quantum annealing device. Finally, I present an algorithm for improving the recovery of ground-state solutions from problems with noise by using thermal fluctuations to infer the correct solution at the Nishimori temperature. While this method has been demonstrated analytically and numerically for trivial ferromagnetic and Gaussian distribu-

tions, a useful metric for more complex Gaussian distributions with added Gaussian noise is unavailable. We show improved recovery of numerical solutions on the chimera graph with a ferromagnetic distribution and added Gaussian noise. Next, I direct my focus to the design of future generations of quantum annealers. The first design is the two-dimensional square-lattice bimodal spin glass with next-nearest ferrromagnetic interactions proposed by Lemke and Campbell claimed to exhibit a finite-temperature spin-glass state for a particular relative strength of the next-nearest to nearest neighbor interactions. Our results from finite-temperature simulations show the system is in a paramagnetic state in the thermodynamic limit, thus not useful for quantum annealing device designs that would benefit from a spin-glass phase transition. The second design is the diluted next-nearest neighbor Ising spin-glass with Gaussian interactions in an attempt to improve the estimation of critical parameter with smaller system sizes by implementing averaging of observables over different graph dilutions. To date, this model has shown no improvement. Finally, I make suggestions for the choice of distributions of interactions that are robust to noise and present a method for using previously unaccessible continuous distributions. I begin with showing the best-case performance of quantum annealing devices. I show results for the resilience, the probability that the ground-state solution has changed due to inherent analog noise in the device, and present strategies for developing robust instance classes. The analog noise is also detrimental to interactions chosen from continuous distributions. Using Gaussian quadratures, I present a method for discretizing continuous distributions to reduce noise effects. Simulations on the D-Wave show that the average residual of the ground-state energy with the true ground-state energy is calculated and shown to be smaller in the case of the discrete distribution.

DEDICATION

To my mother, Ella de la Rosa.

1948 - 2014

ACKNOWLEDGMENTS

First, I would like to thank my research advisor Helmut G. Katzgraber. I am very grateful for all that I have learned from him and his support.

In addition, thanks to the many past and current members of my research group for many fruitful discussions and support throughout my graduate career; Juan Carlos Andresen, Ruben Andrist, Amin Barzegar, Jeff Chancellor, Chao Fang, Darryl Jacob, Richard Lawrence, Ross McDonald, Oliver Melchert, Humberto Muñoz-Bauza, Chris Pattison, Dilina Perera, Wenlong Wang, and especially Zheng Zhu.

I would also like to thank my research collaborators who provided valuable insights during my studies; Firas Hamze, John Machta, Salvatore Mandra, Hidetoshi Nishimori, Kohji Nishimura, Stefan Schnabel, Martin Weigel, and Peter Young.

I am grateful to my committee members, Stephen A. Fulling, Joseph H. Ross, and Kim-Vy H. Tran for their time, effort, and guidance in support of this work.

I would also like to acknowledge the staff at the Department of Physics and Astronomy, the Office of Graduate and Professional Studies, Texas A&M High Performance Research Computing, and Student Counseling Services for all the help I received.

I would like to thank my wife Miyuki Samata for her love and support and my cats Juan Gabriel De La Rosa, Debbie, and Patti for their companionship.

I am very grateful to my family and friends that have supported me along the way. Thank you.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Helmut G. Katzgraber, Kim Vy H. Tran, and Joesph H. Ross of the Department of Physics and Astronomy and Professor(s) Stephen A. Fulling of the Department of Mathematics.

A careful proof reading of this dissertation was performed by Darryl C. Jacob.

- The algorithm presented in Section 4.1 was co-developed with Zheng Zhu and published in 2015 [1].
- The algorithm presented in Section 4.2.3 was co-developed with Zheng Zhu. (Publication in preparation).
- The algorithm presented in Section 4.2.5 was co-developed with Darryl C. Jacob. Data for the benchmarking of this algorithm was provided by Salvatore Mandrà of NASA [2].
- Comparison data in Figure 5.4 was provided by Helmut Katzgraber [3].
- The model in Section 5.2 was simulated by Zheng Zhu. (Publication in preparation).
- The methodology in Section 6.1 was co-developed with Zheng Zhu and published in 2016 [4].

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Graduate Diversity Fellowship from Texas A&M University and the Texas A&M University System Louis Stokes Alliance for Minority Participation (TAMUS LSAMP) Bridge to the Doctorate (BTD) Cohort VIII (2012-2014) Program National Science Foundation Award No. HRD-1249272.

Research efforts were supported in part by the National Science Foundation (Grant No. DMR-1151387). Additional research efforts were supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

I would also like to acknowledge the Texas A&M Department of Physics and Astronomy, Tokyo Institute of Technology, and Physikzentrum Bad Honnef for travel support to conferences and summer schools during my graduate education.

Computational resources were provided by Texas A&M High Performance Research Computing on their Eos, Ada, and Terra clusters. Additional computational resources were provided by Texas Advanced Computing Center (TACC) at the University of Texas at Austin on their Lonestar 4, and Lonestar 5 clusters.

NOMENCLATURE

E	Internal energy
m	Magnetization per spin
N	Number of variables or spins
T	Temperature
w	Distribution of microstates
S	Entropy
k_B	Boltzmann constant
β	Inverse temperature $1/T$
F	Free energy
$\mathcal{Z}$	Partition function
S_i	Ising spin variable
d	Dimension
L	Linear dimension
$\mathcal{H}$	Hamiltonian function
$\mathcal{S}$	Configuration of variables S_i
J_{ij}	Interaction between spin variables S_i and S_j
h_i	Local magnetic field of a spin variable S_i
z	Coordination number
C	Specific heat
M	Magnetization

χ	Susceptibility
ξ	Correlation length
T_c	Curie temperature
$\langle \cdots \rangle$	Ensemble average
$[\cdots]$	Disorder average
t	Deviation from the Curie temperature
g	Binder cumulant
q	Overlap
p_c	Percolation threshold
τ_{eq}	Equilibration time
q_ℓ	Link-overlap
MCMC	Markov Chain Monte Carlo
PT	Parallel Tempering
HCA	Houdayer's Cluster Algorithm
MCS	Monte Carlo Sweeps
SA	Simulated Annealing
QA	Quantum Annealing
ICM	Isoenergetic Cluster Algorithm
DW(2,2X,2000Q)	D-Wave Inc.'s Quantum Annealing Device

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Many research collaborations between government, industry, and academia are investigating an optimization method named *quantum annealing* used by the D-Wave Inc. quantum annealer [5], a commercially available device with a radically new architecture, in order to determine its advantages over conventional optimization algorithms on classical computers [6, 7, 8, 9]. It is believed that quantum fluctuations utilized by quantum annealing machines can help solve these hard problems. However, conclusive evidence of improved performance over classical algorithms used by traditional computers remains elusive [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. The desire for quick and efficient solutions of complex combinatorial optimization problems has inspired the design of new algorithms and novel computing architectures to provide new insights into these difficult optimization problems [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34].

The goal of this dissertation is to present novel research in the study of quantum annealing from a classical, *non-quantum*, perspective by using knowledge gained from the field of statistical physics. In other words, using traditional computers to; understand the nature of complex problems, the physical and theoretical limitations of new quantum annealing devices, and to improve the performance of classical algorithms, we can advance capabilities of both quantum annealing devices and traditional computers to solve difficult optimization problems.

This holistic approach to improving quantum annealing devices begins with an intimate understanding of state-of-the-art classical algorithms in an effort to raise the performance expectations of quantum annealers. The principal goal of a quantum annealer is to find the minimum of a cost function [26]. The cost function is mapped on to a quadratic unconstrained binary optimization problem, the type of problem the D-Wave quantum an-

nealer is designed to optimize. This cost function can represent a plethora of different applications of varying complexity such as number factorization, shortest path through a set of nodes, and job sequencing [35]. A device tailored to solve these problems, such as the D-Wave quantum annealer, could revolutionize optimization efforts.

In order to discover what simulations might reveal the advantages and help overcome the limitations of quantum annealing, we turn our attention to physics. The Ising spin-glass model, from the study of statistical mechanics, can be stated simply. Yet, it is the hardest problem that can be mapped onto a quadratic unconstrained binary optimization problem. Another advantage of studying this physics inspired model is that there is no overhead which reduces the number of variables in the problem when compared to native embedding on the D-Wave quantum annealer. Embedding overhead limits the size of problems one would like to investigate in devices with fixed hardware graphs.

The current generation of quantum annealing devices are limited by analog noise and hardware graph size and connectivity. One might ask, "How will classical simulations help the study of a quantum optimization method?" The goal of classical simulations is two-fold. The first is to benchmark quantum annealing versus other classical optimization methods to determine if there truly is improvement. Second, the current generation of quantum annealing device, the D-Wave quantum annealer, is designed to enable easier embedding of difficult problems, however the current hardware graph design has flaws. Classical simulations can help evaluate and guide graph designs in future devices.

In order to determine if quantum annealing shows improvement over classical algorithms, one must compare it to state-of-the-art classical algorithms. Finding the minimum of a cost function is the first metric of a new optimization method. In addition, for some applications one would like multiple solutions to the problem. With this motivation, I co-developed an algorithm with Zheng Zhu named the *isoenergetic cluster algorithm* which allows a numerical simulation to more efficiently search the phase-space of a problem and,

in some cases, reduce the effort to solve a problem by several orders of magnitude [1]. In addition, this algorithm performs a more fair sampling of problem solutions [36]. Therefore, if a problem has many solutions, they are all found with relatively equal probability. This is not the case in quantum annealing, where some solutions require an exponential amount of effort to be found [9, 2].

This novel algorithm has become the new state-of-the-art for classical simulations in the study of quantum annealing [21, 19, 2, 37]. Due to the general applicability of the algorithm, an implementation of this algorithm went on to win the 2016 MAX-SAT Evaluation [38]. A modified version of this algorithm also allows one to generate new solutions to a problem given a bank of known solutions with effectively zero overhead. This can overcome the problem of some solutions being exponentially suppressed by quantum annealing. In addition, generating additional solutions can assist other optimization algorithms such as machine learning and neural networks that require training on solutions in order to categorize input data.

We know from the study of statistical-mechanics that at a spin-glass phase transition, the energy barriers in this landscape diverge with the problem size. As a result, problems with spin-glass phase transitions are exceptionally difficult for finite-temperature optimization algorithms such as the Monte Carlo method at and below the critical temperature. However, an optimization algorithm such as quantum annealing that takes advantage of quantum fluctuations should be able to tunnel through tall energy barriers and efficiently solve a problem. Thus, it would be advantageous to have a hardware graph that exhibits a phase transition in order to aid benchmarking [10, 15].

With this motivation, we simulated two distinct graphs. First, we studied the two-dimensional square-lattice bimodal spin glass with next-nearest ferromagnetic interactions proposed by Lemke and Campbell [39] which was claimed to exhibit a finite-temperature spin-glass state for a particular relative strength of the next-nearest to nearest neighbor

interaction. An additional advantage of this model is that it has a planar topology that can easily be constructed with current superconducting flux qubits found in the D-Wave quantum annealer. Unfortunately, the model was found to be in a paramagnetic phase at finite-temperatures for large enough system sizes where it is predicted to be a spin glass [40]. Next, we studied the bond-diluted next-nearest-neighbor Ising spin-glass with Gaussian distributed interactions. The goal of this simulation is to determine if averaging over different graphs with site dilutions affects the corrections to scaling. A positive result would imply that performing these graph disorder averages would allow better approximation of critical phenomena with smaller system sizes like those found in the current generation of quantum annealer.

In addition to graph limitations, analog quantum annealing machines will also suffer from analog noise. Both the qubits and the couplers that form the interactions between the qubits, experience this noise. If the noise is large enough, it can cause the device to solve the incorrect problem. We introduced resilience as a measure of the success probability that random field and random bond fluctuations of a problem do not affect the ground-state solution [4]. This allows one to place a classical upper-bound on the performance of an analog quantum annealing device based on the amount of noise and the number of qubits and couplers. With this knowledge we also propose how to generate instance classes, or sets of interactions, that are robust to noise.

Noise also affects the ability to encode continuous distributions onto couplers. As the system size increases, the gap between coupler values becomes infinitely small and very susceptible to noise. To overcome this limitation, we used the method of Gaussian quadratures to discretize continuous distributions. This allows accurate reproduction of thermodynamic variables, however the ground-state manifold becomes degenerate. More importantly, results from the D-Wave showed that the average of the residual of the ground-state energy from the true-ground solution of the problem is lower for the discrete distribution.

This means the D-Wave found energies *closer* to the true solution.

Finally, in an effort to reduce the effects of noise, we introduce algorithms for improving the recovery of ground states from problems with noise by using thermal fluctuations to infer the correct solution at the Nishimori temperature. The method, which was shown analytically by Nishimori [41], proposes a "decoding" step in the simulation where one takes a majority vote per spin, and if the spin was in the correct orientation most of the time during the simulation the correct ground-state could be inferred. We show improvement to solving a trivial ferromagnet with added Gaussian noise.

Using these different approaches, this dissertation contributes in the design of future computing technologies as well as improvement to current state-of-the-art algorithms for use in the simulation of spin-glasses and the optimization of numerically difficult combinatorial optimization problems. Chapter 2 introduces the necessary information regarding complexity of combinatorial optimization problems and their relation to the Ising spin glass. Next, Chapter 3 provides the numerical tools with which to study these hard problems. In Chapter 4, we develop novel algorithms to study spin glasses and their ground-state manifolds that is the new state-of-the-art in classical algorithms in the simulation of spin glasses. In addition, we introduce an algorithm based on analytical results to improve the inference of ground states from problems affected by noise. In Chapter 5, we study the thermodynamic properties of two potential hardware graphs to potentially discover future designs that may yield optimization problems amenable to quantum annealing. Finally, in Chapter 6, we focus on the problem of noise in current quantum annealing hardware and describe general methods to calculate best case success probability as well as encode previously unaccessible continuous distributions.

2. PRELIMINARIES

The first commercial quantum annealing device that attempts to exploit the unique power of quantum fluctuations is the D-Wave quantum annealing machine [5]. This device is designed to solve Quadratic Unconstrained Binary Optimization problems (QUBOs) [35], such as finding the ground state of a disordered Ising spin-glass Hamiltonian [26]. Furthermore, though the Ising spin glass has its roots in statistical mechanics, many complex problems can be mapped onto it. This chapter introduces the necessary information about the complexity of these problems and their relation to spin glasses along with an overview of spin-glass physics.

2.1 Complexity

Typical goals of an optimization problem are to compute an observable, for instance the energy or magnetization of a spin glass, or minimize the cost function of a problem. Some examples of optimization problems are satisfiability, number partitioning, vertex cover, and traveling salesman problems [35]. In addition to minimizing the cost function, one would like to do so quickly.

As a general rule of thumb, the time complexity of an algorithm is a function of the number of variables. In a spin glass with N spins, finding a solution amounts to searching for a configuration among 2^N others that minimizes the energy. A *desirable* algorithm is one that requires a number of calculations bounded by a sub-exponential function of the size of the problem. For this type of problem, if the number of calculations grows exponentially with the size of the problem, then the problem should be considered intractable.

2.1.1 Complexity Classes

There is a veritable menagerie of complexity classes. However, I will restrict myself to discussing the three most relevant.

P, or polynomial complexity, is the set of problems that can be solved deterministically in an amount of time that is polynomial in the size of the input. For example, $\tau \propto N^2$ in the case of Euclid's algorithm to calculate the greatest common divisor of two or more integers where N is the average number of digits in those integers. NP or non-deterministic polynomial-time complexity is the class of all decision problems for which a solution can be verified as correct in polynomial time.

There exists a subclass of NP problems such that if one can solve one of these problems, then one can solve all NP problems [42]. This class of problems is named "NP-complete" and it contains the hardest problems in NP. It includes many 'classical' problems in combinatorics such as the traveling salesman problem, graph coloring problem, and all problems in the class that have been shown to be equivalent in the sense that if one problem is tractable, they all are [43]. In addition to their hardness, NP-complete problems' time complexity scales worse than any polynomial, i.e., $\tau \propto 2^N$ as in the non-planar Ising spin glass.

One example of a problem where a quantum algorithm may be useful is integer factorization, where the best published running time of the general number field sieve algorithm is on the order of $\exp(\frac{64}{9}b(\log b)^2)^{(1/3)}$ for b-bit numbers [44]. However, for a programmable quantum computer, there exists Shor's algorithm which has a runtime on the order of b^3 thus solving the problem in polynomial time [45]. This is just one example of the significant implications if quantum computation is possible.

All NP-complete class problems can be mapped onto the Ising spin-glass. For this reason, we turn our attention to the statistical properties of spin glasses to introduce and

better understand this fundamental example of an optimization problem. In addition, this is the native problem that current quantum annealing machines, such as the D-Wave quantum annealing device, are designed to solve. Thus, if one can improve the optimization of spin glasses, this result will extend to many other combinatorial optimization problems. Optimization methods relevant to this dissertation are discussed in Ch. 3.

2.2 The Ising Spin Glass

As mentioned in the previous section, the Ising spin glass is a QUBO problem, the type of problem the current generation of quantum annealers are designed to solve. The Ising spin glass is one of the simplest models to study among systems with many interacting elements. In the special case of the ferromagnetic Ising model [46], it is also the only model that is analytically solvable, while exhibiting nontrivial critical behavior in the form of a phase transition [47]. Finally, it can be extended in such a plethora of ways that it is often referred to as the "fruit fly" of statistical mechanics [48, 49, 50]. As such, it plays an integral role in statistical physics and in our current study of benchmarking quantum annealing.

2.2.1 Definition

The Ising spin glass is a collection of vertexes $\mathcal{V}$ of a simple undirected graph $\mathcal{G}$ with edges $\mathcal{E}$. At each vertex is a spin variable S_i $(i = 1, ..., N)$ with $S_i \in \{\pm 1\}$. A positive value of S_i is occasionally referred to as *spin up*, and a negative value is referred to as *spin down*. A set of values $\{S_i\}$ specifies the configuration of the system, $\mathcal{S}$. For any two sites connected by an edge i, j, there is an interaction J_{ij}. The energy of a given configuration is

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J_{ij} S_i S_j - \sum_{i=1}^{N} h_i S_i \qquad \text{with} \qquad S_i \in \{\pm 1\}, \tag{2.1}$$

with $\langle i, j \rangle$ representing a sum over the edges $\mathcal{E}$, and h_i, the value of a magnetic field if present. Interactions are symmetric, in other words, $J_{ij} = J_{ji}$. Thus there are $zN/2$ terms in the sum of Eq. 2.1,where z is the number of nearest neighbors or coordination number, for example, $z = 4$ for the two-dimensional square lattice. For the remainder of this dissertation I will focus on Ising spin glasses that are connected by nearest-neighbor and next-nearest neighbor interactions.

2.2.2 Properties

The case $J_{ij} > 0$ corresponds to ferromagnetism, an interesting phenomenon in solid state physics where some fraction of spins become polarized in the same direction, resulting in a net magnetic moment. According to the Hamiltonian in Eq. (2.1), spins that are aligned will minimize the energy. Similarly, $J_{ij} < 0$ corresponds to antiferromagnetism, where opposing neighbor spins minimize the energy.

The Ising spin glass is the bond-disordered version of the Ising ferromagnet and can be understood as a collection of random spin-spin interactions J_{ij} [51]. The choice of interactions depends on the type of problem. Typically, the disorder is either chosen from a bimodal distribution or a Gaussian distribution with zero mean and standard deviation unity. In all cases, when one wants to evaluate a physical quantity from the Hamiltonian, one begins with a given fixed or *quenched* set of J_{ij}, generated by a probability distribution.

$$P(J_{ij}) = p\delta(J_{ij} - J) + (1 - p)\delta(J_{ij} + J) \tag{2.2}$$

$$P(J_{ij}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ - \frac{(J_{ij} - \mu)^2}{2\sigma^2} \right\} \tag{2.3}$$

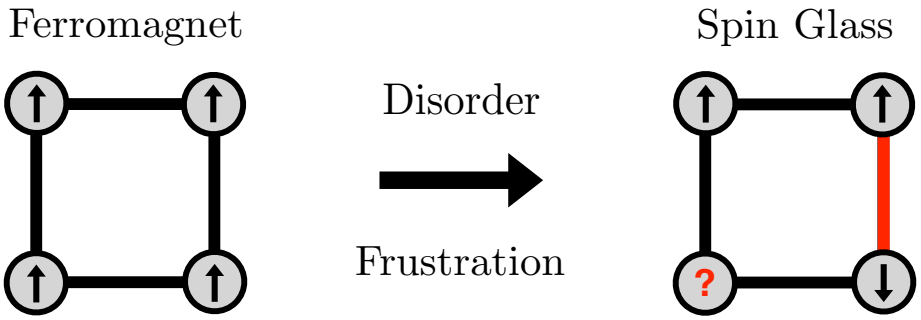

Figure 2.1: Illustration of disorder and frustration in a spin glass. Interactions in black are ferromagnetic and the interaction in red is antiferromagnetic. In a ferromagnet, the energy is minimized when all spins are aligned in the same direction. In the spin glass plaquette illustrated, no combination of spin directions can satisfy all interactions and the system is considered to be frustrated.

Frustration, and hence glassy order, occurs when neighboring spins have combinations of ferromagnetic and antiferromagnetic interactions such that the center spin has no preferred orientation that minimizes the energy which greatly increases the complexity of a problem. For example, in the spin glass shown in Fig. 2.1, no assignment of the bottom-left spin can satisfy the neighboring interactions.

2.2.3 Order Parameters

The general recipe of statistical mechanics is to calculate the thermal average $\langle \cdots \rangle$ of a physical quantity $\mathcal{O}$ using the Gibbs distribution where $\mathcal{S}$ is a configuration of variables and $\mathcal{Z}$ is the partition function.

$$\langle \mathcal{O} \rangle = \sum_{\mathcal{S}} \mathcal{O}(\mathcal{S}) \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\mathcal{S})} \qquad \text{with} \qquad \beta = \frac{1}{k_B T} \tag{2.4}$$

The usual quantity used to measure the macroscopic properties of the Ising model with ferromagnetic ($J > 0$) interactions is the magnetization.

$$m = \frac{1}{N}\left\langle \sum_{i=1}^{N} S_i \right\rangle \tag{2.5}$$

Magnetization is a typical example of an order parameter that measures the overall ordering in a macroscopic system. If equal numbers of spin up $S_i = 1$ and spin down $S_i = -1$ exist, the magnetization vanishes and no uniform ordered state exists. At low energies, according to the Gibbs distribution, low energy states are preferred. The low-energy states with $h = 0$ have all spins in the same directions. Thus, the magnetization m is likely very close to 1 or -1. As the temperature increases, states with various energies appear with similar probabilities. Here, S_i changes frequently so that the system is disordered with vanishing magnetization.

In the case of random J_{ij}, the average magnetization $\langle m \rangle$ vanishes everywhere. Thus, a new order parameter, overlap (q), is required. In addition to thermal averaging to measure an observable, a disorder average denoted by $[\cdots]$, is required for spin glasses. This is due to the inability to simulate an infinite lattice and requires us to investigate small subsystems with different interactions chosen from the same disorder distribution.

$$q = \left[\left\langle \sum_{i}^{N} S_i^{\alpha} S_i^{\beta} \right\rangle \right], \tag{2.6}$$

where α and β denote two different replicas of the same system. Two replicas are required because in a spin glass, the frozen state and the unordered state are indistinguishable. By simulating two replicas one can compare the two configurations to determine if the system is frozen and $\langle q \rangle > 0$ or the system is in the unordered phase and $\langle q \rangle = 0$.

One advantage of studying spin glasses is that the energy landscape is much more

complex. Instead of a singular valley, for a simple $J_{ij} > 0$ Ising model, the new 2^N dimensional landscape is much more "rugged." This additional complexity creates very challenging problems for optimization algorithms as shown in Ch. 3.

In the ferromagnetic Ising model, there is a critical temperature or Curie temperature, T_c for which $|m| \neq 0$ for $T < T_c$ and, $m = 0$ for $T > T_c$. This phenomenon is called a phase transition. The two-dimensional Ising model is one nontrivial example of a phase transition that can be treated analytically [47]. The Ising spin glass can also have phase transitions as we will see in Sec. 5.2. However, for simplicity we introduce the phase transitions for the ferromagnetic Ising model first.

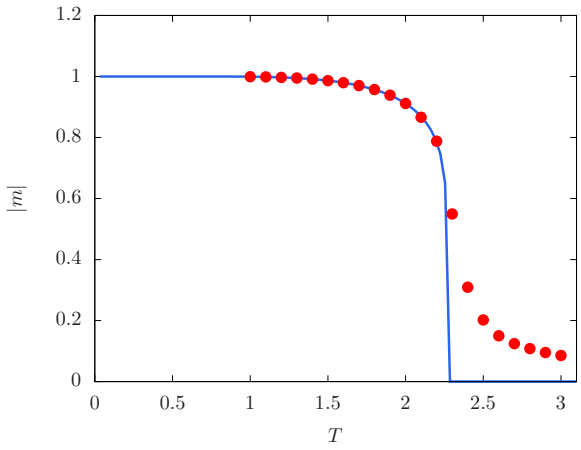

Figure 2.2: Numerical simulation of magnetization per spin (circles) as a function of temperature T for the two-dimensional Ising model with $N = 1024$ spins. The solid line is the analytical value of the magnetization for an infinite square lattice. Due to finite size effects in the simulation, the phase transition of the magnetization is spread out instead of becoming zero above the critical temperature.

2.2.4 Phase Transitions

A phase transition occurs when there is a singularity in the free energy or one of its derivatives [52]. As described earlier, the order parameter in the Ising model is the mag-

netization in a ferromagnetic system. The net magnetization's direction is spontaneously chosen when the system cools below the Curie temperature and the phase changes from paramagnetic to ferromagnetic as shown in Fig. 2.2.

There are two types of phase transitions; first-order and second-order phase transitions. A first-order transition is typically one with latent heat, while a second-order transition, or "continuous" phase transition is one in which the second derivative of a thermodynamic variable diverges. In the case of the two-dimensional ferromagnetic Ising model, at the critical temperature, the magnetic susceptibility $\chi_m = \partial M/\partial H$ will typically diverge as shown in Fig. 2.3.

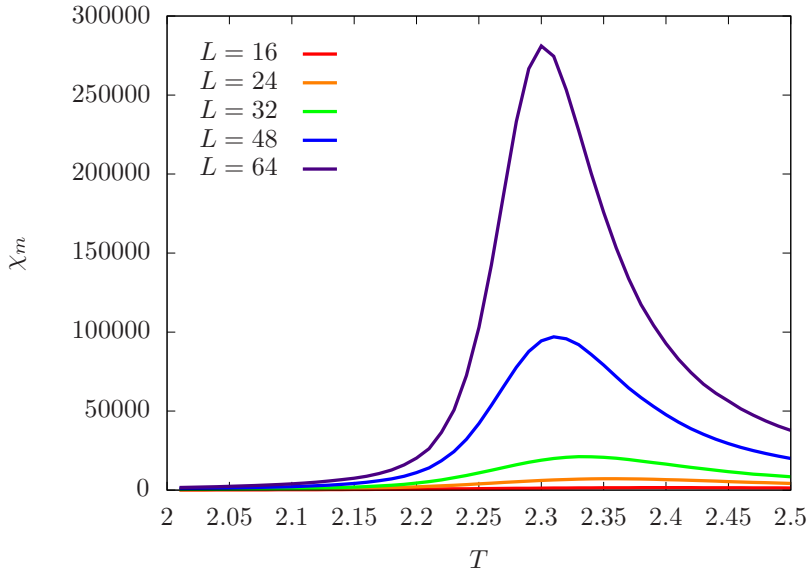

Figure 2.3: Magnetic susceptibility as a function of temperature for the two-dimensional Ising model with $N = L^2$ spins. As the system size increases χ_m diverges which is a signal of a phase transition.

In the case of the two-dimensional Ising model with periodic boundary conditions, in the limit of $N \to \infty$, the critical temperature and spontaneous magnetization for $T < T_c$

can be calculated analytically [47].

$$T_c = 2/\ln(1+\sqrt{2}) \tag{2.7}$$

$$M = \left[1 - \sinh^{-4}(2\beta J)\right]^{1/8} \qquad \text{with} \qquad \beta = \frac{1}{T} \tag{2.8}$$

Here, the derivative of the magnetization M diverges at T_c indicating the phase transition from paramagnetic to ferromagnetic.

A phase transition alone does not characterize a system. T_c depends sensitively on the details of the interactions J_{ij}. In order to understand the nature of a statistical system, it is required to understand the critical exponents of the system.

2.2.5 Critical Exponents

It is important to understand the divergences in magnetic susceptibility and specific heat. This will be used in Sec. 5.1 to determine if corrections to scaling can be improved. First, we define a measure of the deviation in temperature from the critical temperature T_c.

$$t = \frac{T - T_c}{T_c} \tag{2.9}$$

Then we define the critical exponent associated with a function f.

$$\lambda = \lim_{t \to 0} \frac{\ln|f(t)|}{\ln|t|} \qquad \text{or} \qquad f(t) \sim |t|^{\lambda} \tag{2.10}$$

This implies that close to the transition, the quantity f is dominated by a non-analytic part $f(t) \sim t^{\lambda}$ for $t \to 0$. The importance in the critical exponents lies in the fact that while T_c depends on inter-atomic interactions, the critical exponents are universal. A list of critical exponents is given in Table 2.1. These critical exponents depend on spacial dimension d and order parameter symmetry.

Knowledge of the critical exponent of a simple model system can be used to obtain critical exponents for all systems in a universality class. Critical exponents are related by scaling relations. In most cases, only two exponents are necessary to fully characterize the behavior of a model. If one determines the location of the critical temperature T_c and two independent critical exponents, one can deduce the universality class of the model and calculate all other critical exponents.

Zero-field specific heat	$C \sim \|t\|^{-\alpha}$
Zero-field magnetization	$M \sim (-t)^{\beta}$
Zero-field isothermal susceptibility	$\chi_T \sim \|t\|^{-\gamma}$
Correlation length	$\xi \sim \|t\|^{-\nu}$

Table 2.1: Definitions of the most commonly used critical exponents for a magnetic system. Adapted from Yeomans [52].

Unfortunately, it is difficult to obtain accurate critical exponents because there are significant corrections to scaling. There exist long equilibration times in Monte Carlo simulations that limit the available system sizes, and all quantities need to be averaged over many disorder realizations in order to reduce error. A major problem with reducing error bars in critical exponents is the presence of corrections to finite size scaling, which means that the scaling expressions used to determine exponents do not work well for small (finite) system sizes [3].

2.2.6 Finite Size Scaling Analysis

Suppose we would like to determine the critical exponents of an infinite system by simulating finite lattices. As shown in Fig. 2.2, when the system is not infinitely large, the critical behavior is smeared. However, one can show that the non-analytic part of

a given observable can be described by a finite-size scaling form [53]. The finite-size magnetization of an Ising model with L^d spins, where L is the linear dimension, close to the transition and for large L, is given by

$$\langle m_L \rangle \sim L^{\beta/\nu} \tilde{M}[L^{1/\nu}(T - T_c)]. \tag{2.11}$$

$\tilde{M}$ is an unknown scaling function. According to Eq. (2.11), the data for $\langle m_L \rangle$ should align in the large-L limit at $T = T_c$ if we can determine β and ν correctly. There are methods to approximate these but a simpler method is to use combined quantities that are dimensionless, such as the Binder ratio or the Binder cumulant [54].

$$g_m = \frac{1}{2}\left[3 - \frac{\langle m^4 \rangle}{\langle m^2 \rangle^2}\right] \sim \tilde{G}[L^{1/\nu}(T - T_c)] \tag{2.12}$$

The Binder ratio is a dimensionless quantity. Thus, data for multiple system sizes N will cross at a transition. By approximating a correct value for the critical exponent ν, the data falls onto a universal curve as shown in Fig. 2.4 and it is now possible to estimate both T_c and ν.

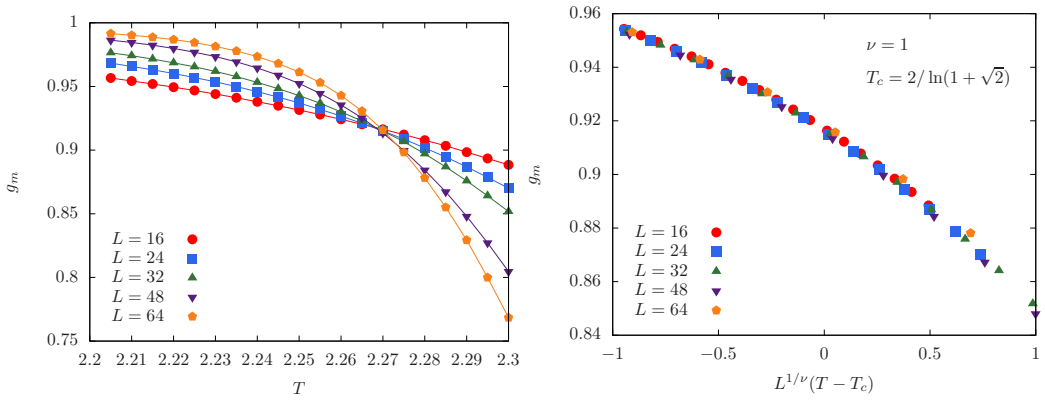

Figure 2.4: Left: Binder ratio as a function of temperature for the two-dimensional square lattice $J = 1$ Ising model. The data approximately cross near the paramagnetic to ferromagnetic phase transition at $T_c \approx 2.269$. Right: Finite-size scaling analysis with the known values of ν and T_c. The data fall on a universal curve.

3. NUMERICAL METHODS

As described in the previous section, the Ising spin glass is a QUBO problem native to the current generation of quantum annealing hardware. There are many optimization methods to choose from when trying to find the minimum of the Ising spin-glass Hamiltonian. This chapter provides a review of state-of-the-art numerical methods and relevant optimization algorithms.

In the study of spin glasses from a physics perspective, one typically would like to perform finite-temperature simulations in order to estimate observables in equilibrium with a heat bath for which the partition function is defined. On the contrary, in optimization problems, one simply requires the minimum of a cost function as quickly as possible. Fortunately, our algorithms for studying physical quantities can also be modified to find a solution quickly by simulating low-temperatures where the probability of being in the ground state is most likely.

Calculating the sum over 2^N terms of a partition function is usually very difficult. For a simple 32×32 square lattice, there are approximately 10^{308} possible configurations of spins. If one considers there are approximately 10^{80} atoms in the universe, it quickly becomes apparent that even small system sizes are intractable for enumerative methods.

One approximate approach to overcome this limitation is the Monte Carlo method, which is well suited for computers. The Monte Carlo method is particularly useful in generating draws from a probability distribution and for simulating systems such as disordered materials and interacting particle systems.

We wish to compute the average of an observable $\mathcal{O}$.

$$\langle \mathcal{O} \rangle = \frac{\sum_S \mathcal{O}(S) e^{-\mathcal{H}(S)/k_B T}}{\sum_S e^{-\mathcal{H}(S)/k_B T}} \tag{3.1}$$

If we extend this to a distribution $\mathcal{P}$ and use the Boltzmann distribution then we arrive at an estimate for $\mathcal{O}$ where the states are selected according to a Boltzmann distribution.

$$\langle \mathcal{O} \rangle = \frac{\sum_S [\mathcal{O}(S)/\mathcal{P}(S)] e^{-\mathcal{H}(S)/k_B T}}{\sum_S [1/\mathcal{P}(S)] e^{-\mathcal{H}(S)/k_B T}} \tag{3.2}$$

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_i^N \mathcal{O}(S_i) \tag{3.3}$$

If we can sample $\mathcal{O}$ from the Boltzmann distribution using the Metropolis algorithm then $e^{-\mathcal{H}(S)/k_B T}/\mathcal{P}(\mathcal{S}) = 1$ and Eq. (3.2) reduces to Eq. (3.3).

3.1 Metropolis Algorithm

In simple Monte Carlo sampling, a random state (configuration) $\mathcal{S}$ is generated and observables are measured. However, this can be computationally inefficient. Instead of a random search, a Markov process is used to generate a new state from the previous state. This is accomplished by using the Metropolis method to sample the Boltzmann distribution. It is possible to choose the transition probability $\mathcal{T}(\mathcal{S}_n \to \mathcal{S}_{n+1})$ such that it occurs with a probability given by the Boltzmann distribution, $\mathcal{P}_{\rm eq}(\mathcal{S}) = \mathcal{Z}^{-1} \exp[-\mathcal{H}(\mathcal{S})/k_B T]$.

In the Markov process, the state $\mathcal{S}$ occurs with probability $\mathcal{P}_k(\mathcal{S})$ at the kth time step described by the *master equation*, with the goal as $k \to \infty$, the probability $\mathcal{P}_k \to \mathcal{P}_{\rm eq}$.

$$\mathcal{P}_{k+1}(\mathcal{S}_{n+1}) = \mathcal{P}_k(\mathcal{S}_n) + \sum_{\mathcal{S}_{n+1}} [\mathcal{T}(\mathcal{S}_{n+1} \to \mathcal{S}_n)\mathcal{P}_k(\mathcal{S}_{n+1}) - \mathcal{T}(\mathcal{S}_n \to \mathcal{S}_{n+1})\mathcal{P}_k(\mathcal{S}_n)] \tag{3.4}$$

Transition probabilities $\mathcal{T}$ can be chosen so that $\mathcal{P}_k = \mathcal{P}_{\rm eq}$. This causes terms in the sum to evaluate to zero and implies that, for all $\mathcal{S}_n$ and $\mathcal{S}_{n+1}$, the detailed balance condition holds.

$$\mathcal{T}(\mathcal{S}_{n+1} \to \mathcal{S}_n)\mathcal{P}_{\rm eq}(\mathcal{S}_{n+1}) = \mathcal{T}(\mathcal{S}_n \to \mathcal{S}_{n+1})\mathcal{P}_{\rm eq}(\mathcal{S}_n) \tag{3.5}$$

Detailed balance is important because this ensures that the process is reversible. Further-

more, when the system has assumed the equilibrium probabilities, the ratio of the transition probabilities only depends on the change in energy

$$\Delta\mathcal{H}(\mathcal{S}_n, \mathcal{S}_{n+1}) = \mathcal{H}(\mathcal{S}_{n+1}) - \mathcal{H}(\mathcal{S}_n), \tag{3.6}$$

$$\frac{\mathcal{T}(\mathcal{S}_n \to \mathcal{S}_{n+1})}{\mathcal{T}(\mathcal{S}_{n+1} \to \mathcal{S}_n)} = \exp[-(\mathcal{H}(\mathcal{S}_{n+1}) - \mathcal{H}(\mathcal{S}_n)/k_B T] = \exp[-\Delta\mathcal{H}(\mathcal{S}_n, \mathcal{S}_{n+1})/k_b T]. \tag{3.7}$$

The Metropolis algorithm gives us a choice of $\mathcal{T}$ that satisfies Eq. (3.7). If the energy is minimized, the new configuration is accepted with probability 1. However, if the energy is not minimized, the new configuration is accepted with a probability that depends on the change in energy, $\Delta\mathcal{H}$, and temperature T.

$$\mathcal{T}(\mathcal{S}_n \to \mathcal{S}_{n+1}) = \begin{cases} 1, & \text{if } \Delta\mathcal{H} \leq 0 \\ \exp[-\Delta\mathcal{H}(\mathcal{S}_n, \mathcal{S}_{n+1})\beta], & \text{if } \Delta\mathcal{H} \geq 0 \end{cases} \tag{3.8}$$

For Ising models with ferromagnetic interactions, i.e., $J_{ij} = 1$, this works quite well. However, at the transition temperature, T_c, the method fails due to critical slowing down of dynamics. When disorder is added, simple Monte Carlo is insufficient. The energy landscape transforms from one smooth valley into a very jagged and mountainous landscape with many metastable states. When attempting to flip a single spin, acceptance probabilities are often exponentially small, rendering the method inefficient.

The Metropolis algorithm does not take into consideration that sometimes even a simple spin flip can produce a large change in the energy $\Delta\mathcal{H}$ of the system. As shown in Eq. (3.8), a large $\Delta\mathcal{H}$ leads to an exponentially small transition probability and the simulation will stall. Complex systems such as spin glasses and neural networks have the features of a rough energy landscape, i.e., different states in phase space are separated by large energy gaps. For these systems at low temperatures, simple Monte Carlo methods

diverge. Therefore, optimized sampling techniques are required. One can attempt to improve the local updating technique by introducing artificial statistical ensembles such that the time to climb over energy barriers or tunnel through is reduced.

3.1.1 Parallel Tempering

In order to efficiently overcome barriers, an acceleration of dynamics must be added to Markov chain Monte Carlo (MCMC). The present research uses the Exchange Monte Carlo method, also known as parallel tempering (PT), developed by Hukushima and Nemoto [55]. If the change in energy when attempting to flip a spin is too large or the temperature T is too low, the probability to accept a Monte Carlo move and escape a metastable state exponentially decreases. Parallel tempering accelerates thermalization by allowing configurations to be heated and cooled throughout the simulation.As shown in Fig. 3.1, as the simulation moves up in temperature space, the probability to accept Metropolis moves increases, allowing the configuration to escape local minima in the energy landscape.

In this method, multiple non-interacting replicas are simulated at different temperatures independently and simultaneously as a canonical ensemble. After a fixed number of Monte Carlo sweeps, typically one sweep, two replicas at neighboring temperatures are exchanged with a Metropolis acceptance probability:

$$\mathcal{P}(\mathcal{S}_i \to \mathcal{S}_j) = \min[1, \exp\left(-\Delta\beta\Delta\mathcal{H}\right)] \tag{3.9}$$

The replica exchange is restricted to neighboring temperatures because the acceptance ratio decreases exponentially with $\Delta\beta = \beta_{n+1} - \beta_n$. A careful design of the temperature set being used is needed for the method to be efficient. However, many recipes are readily available [56]. A replica will perform a random walk over temperature space to high temperatures, where equilibration is rapid, and back to low temperatures, where there is critical slowing down. Furthermore, the simulation of additional replicas to use paral-

lel tempering does not increase overhead due to the desire to simulate observables as a function of temperature.

Parallel tempering can be combined with other types of spin updates to improve thermalization time. It has also been used successfully in other fields of physics to simulate biomolecules, determine x-ray structure, and study molecular dynamics [57].

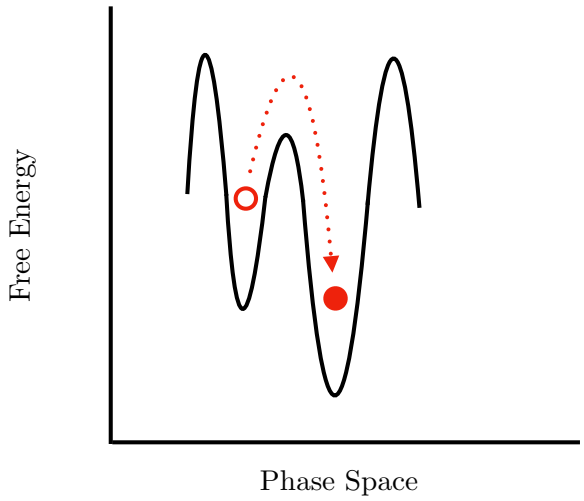

Figure 3.1: Illustration of parallel tempering in a rugged energy landscape in a spin glass due to disorder and frustration. A Metropolis move from the initial (open circle) to final state (closed circle) is unlikely if the size of the barrier ΔE is too large or if the temperature is too low. A simple Monte Carlo simulation will be stuck in a local minimum. In the illustration, parallel tempering (dotted line) allows the simulation to increase in energy as the temperature increases, then cools to allow the simulation to reach a new minimum.

3.1.2 Houdayer's Cluster Algorithm

Despite the improvement in thermalization gained from parallel tempering, it is still a local update algorithm. One method to improve thermalization further is global updates to the spin configuration, or cluster moves. Cluster algorithms offer an additional method by which to increase mixing of configurations in spin glasses and speed up thermalization

times at low temperatures or near phase transitions where simple Monte Carlo has difficulties due to small spin-flip probabilities depending on temperature or large change in energy respectively.

There are several well known cluster algorithms in the study of spin systems. For the simple ferromagnetic Ising model, there exists the Wolff cluster algorithm that builds a cluster according to a probability and always flips the cluster thus "teleporting" the system to a different region of phase space [58]. For spin-glasses there is the Swensden-Wang algorithm which reduces to parallel tempering from the previous subsection in dimensions larger than two [59]. However, the aforementioned cluster algorithms, based on ideas from percolation theory are limited to two-dimensions where the probability to have a cluster that spans the system, or *percolate*, is below the site percolation threshold. Perhaps the most useful cluster algorithm for the study of spin-glasses is Houdayer's cluster algorithm [60].

In this section I review Houdayer's cluster algorithm which in the following chapter will be the basis of a novel cluster algorithm I have co-developed.

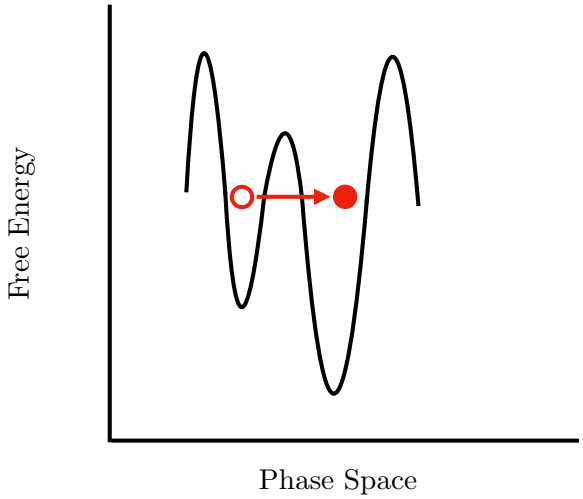

Figure 3.2: Illustration of a cluster move through the rugged energy landscape in a spin glass due to disorder and frustration. Cluster moves introduce large rearrangements of spins in one time step in order to "teleport" through barriers in energy that would be difficult to overcome with simple Monte Carlo or parallel tempering.

In parallel tempering, configurations can move up and down in the energy landscape by changing the simulation temperature dynamically. However, large rearrangements of spins would allow the simulation to move through phase space at a much greater rate as shown in Fig. 3.2. There are multiple cluster algorithms for dimension less than or equal to two [59, 58, 60]. And yet, a cluster algorithm for Ising spin glasses with space dimension greater than two is elusive. Jerome Houdayer proposed a cluster algorithm for two-dimensional spin glasses [60]. By adding Houdayer's cluster moves, configurations can also "teleport horizontally" through the landscape while keeping the sum of energies *constant*. This means the algorithm is rejection free, i.e., extremely efficient because there is neither a need to calculate a probability of flipping the spin nor the need to generate and compare it to random numbers. Since the system keeps the energy constant, configurations randomize efficiently and barriers are overcome easily for two-dimensional systems.

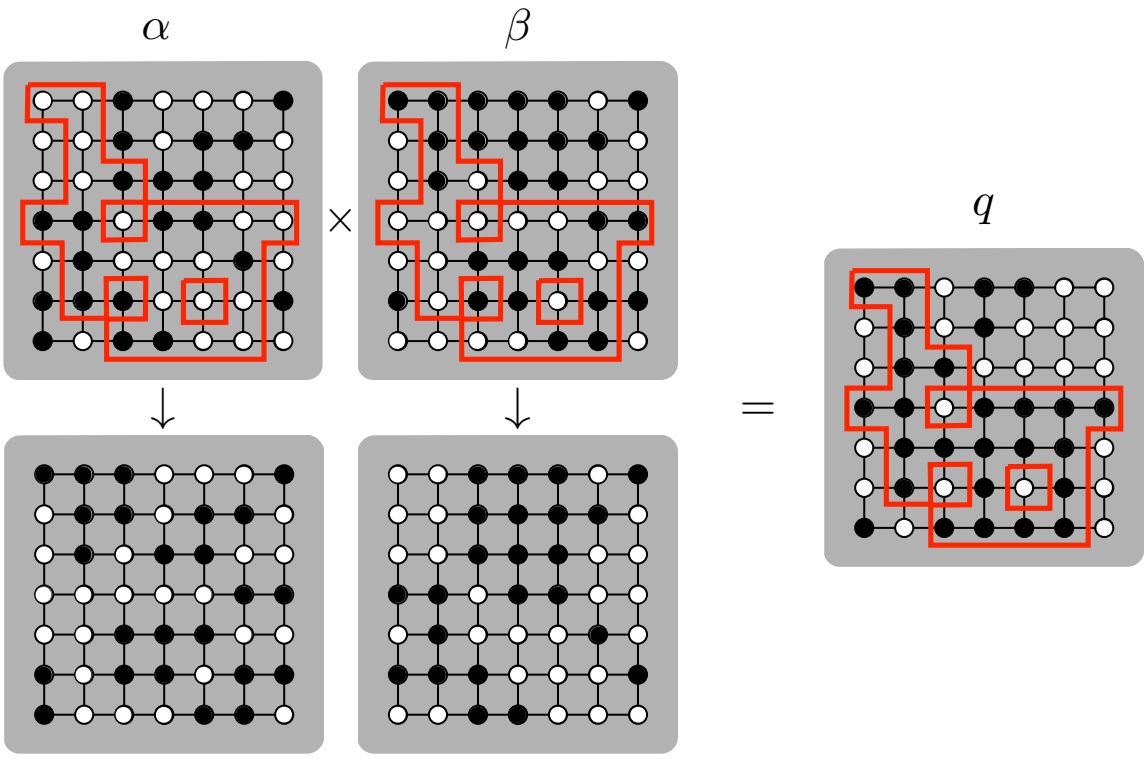

Figure 3.3: Illustration of a Houdayer cluster move. Spins with $S_i = 1$ are drawn in white and $S_i = -1$ in black. The overlap of two replicas α and β at the same temperature is calculated. A cluster of spins with $q_i = -1$ is chosen (in red), and spins that make the cluster are flipped in the α and β replicas resulting in two new configurations.

Consider a system of two independent replicas α and β at the same temperature. Within this context a replica represents a copy of a system with the same disorder but a different Markov chain. The local overlap at site i between the two replicas is defined by $q_i = S_i^\alpha S_i^\beta$. Within replica space, two domains can occur: one with $q_i = 1$ and one with $q_i = -1$. The clusters are defined as the connected parts of these domains. The cluster step begins by choosing one site at random for which $q_i = -1$. Neighbors are added with probability 1 if and only if for a neighbor of spin i, $q_{\mathrm{nb}(i)} = -1$, until no more spins can be added to the connected backbone of the cluster. Spins within both replicas that correspond to cluster members are then flipped. An illustration of a Houdayer cluster move is shown in

Fig. 3.3. Note that this does not change the sum of the energies, i.e., $\Delta E = \mathcal{H}(\mathcal{S}^\alpha) + \mathcal{H}(\mathcal{S}^\beta) = 0$. Since the replicas are at the same temperature, the cluster move is accepted with probability 1. To enforce ergodicity, the cluster move is combined with a standard single-spin flip Monte Carlo move with the Metropolis acceptance probability. Using additional replicas allows for much faster thermalization because the replicas are mixed very quickly. Following the cluster update, a parallel tempering move is performed. The resulting algorithm is very efficient because it is able to explore the energy landscape both "vertically", via parallel tempering, and "horizontally", via Houdayer cluster moves.

To summarize, one simulation step of the simulation consists of the following steps:

- Perform one Monte Carlo sweep (N Metropolis updates) for each replica.
- Perform one Houdayer cluster move.
- Perform one parallel tempering update for a pair of neighboring temperatures.

Note that the last step is not necessary; however, the combination of the Houdayer cluster moves and parallel tempering updates improves thermalization considerably and represents the standard *modus operandi*.

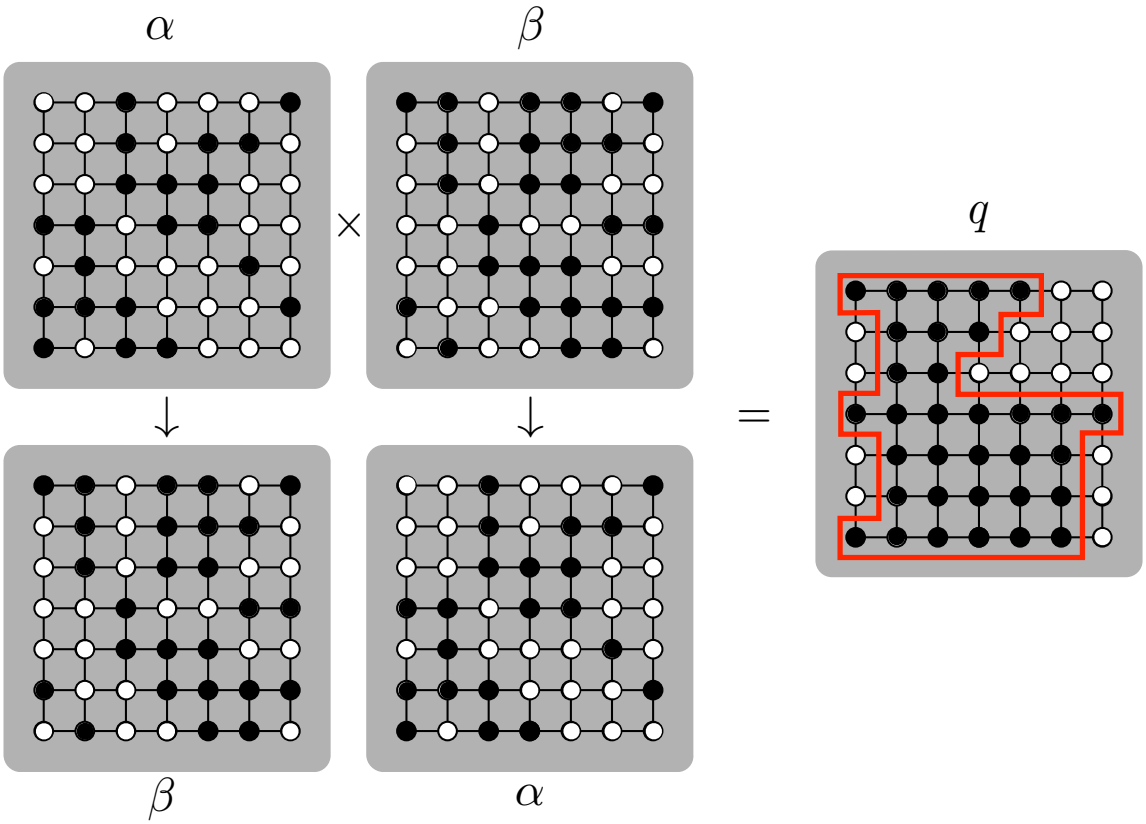

Figure 3.4: Illustration of a failed Houdayer cluster move. Spins with $S_i = 1$ are drawn in white and $S_i = -1$ in black. The overlap of two replicas α and β at the same temperature is calculated. A cluster of spins with $q_i = -1$ is chosen (in red), and spins that make the cluster are flipped in the α and β replicas resulting in no new configurations. Due to the percolating cluster that spans the system, performing a cluster move is equivalent to exchanging the two replicas.

In principle, one could expect cluster algorithms to only work below the percolation threshold where clusters do not span the whole system, i.e., flipping a cluster only globally affects the system but does not efficiently randomize the spin configurations. In fact, Houdayer claims the problem is encountered as soon as the site percolation threshold is less than 0.5 [60]. This causes two cluster to form with $q_i = 1$ and $q_i = -1$. Flipping one of the large clusters becomes equivalent to exchanging the two configurations and no efficient randomization occurs as shown in Fig. 3.4. In two space dimensions, the percolation

threshold is $p_c \approx 0.59$ [61]. As such, the method works efficiently in this case. However, as we shall see, the percolation threshold for three space dimensions ($p_c \approx 0.3116$) [62] and the D-Wave Two quantum annealing machine's Chimera [$p_c \approx 0.3866$] [1] topologies are below 0.5 meaning the clusters are likely to percolate. In Sec. 4.1, we propose an effective solution.

3.2 Equilibration Techniques

We wish to sample a distribution in thermal equilibrium. To do this, one should monitor all observables as a function of time, $\mathcal{O}(\tau)$. The time it takes for $\mathcal{O}(\tau)$ to be constant is the equilibration time. Because the initial configuration is random, some number of algorithm steps are required to reach an appropriate estimate of any observable. The equilibration time has several important properties. τ_{eq} increases with system size, increases with decreasing temperature, and is typically measured in Monte Carlo sweeps (MCS) or N spin update attempts.

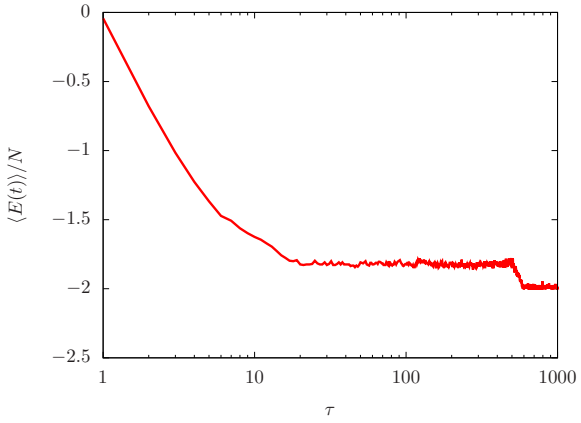

Figure 3.5: Energy per spin as a function of time in Monte Carlo sweeps for the two-dimensional Ising model with $N = 1024$ spins via simple Metropolis Monte Carlo at $T \sim 0.1T_c$. The solution of the two-dimensional ferromagnetic Ising model is $E/N = -2$. After approximately 600 sweeps the system fluctuates near the ground state.

In Fig. 3.5, the simulation begins in a random configuration that has an energy close to zero. Very quickly, the simulation reaches a metastable state where two domain of spins exist. The boundary between these two domains or, *domain wall*, increases the energy of the system. However, after some amount of time, the system is able to push out the domain wall and, in this case, fluctuate around the ground state energy.

In the particular case of the Ising spin glass with normally-distributed disorder, one can use an exact relationship between the energy and link overlap q_ℓ.

$$q_\ell = \frac{1}{dN} \sum_{\langle i,j \rangle} S_i^\alpha S_j^\alpha S_i^\beta S_j^\beta \tag{3.10}$$

The link overlap measures the average length of the boundary of a flipped domain. The internal energy per spin u is given by

$$u = -\frac{1}{N} \sum_{\langle i,j \rangle} [J_{ij} \langle S_i S_j \rangle] \tag{3.11}$$

It is possible to perform an integration by parts of J_{ij} to relate u to the link overlap [63].

$$[\langle q_\ell \rangle] = 1 + \frac{Tu}{d} \tag{3.12}$$

By beginning the simulation with a random spin configuration, the measure of q_ℓ will be small in magnitude and the initial energy will will be higher as it would not be in thermal equilibrium. Thus the two sides of Eq. (3.12) will approach equilibrium from opposite directions, as shown in Fig. 3.6.

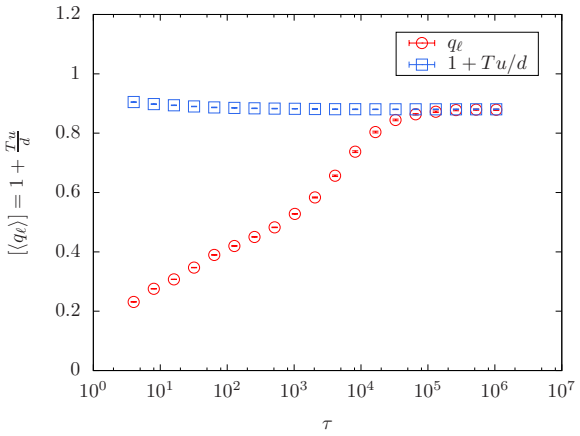

Figure 3.6: A measure of low-temperature, $T \sim 0.2T_c$, equilibration for the two-dimensional Ising spin-glass with $N = 1024$ spins and Gaussian distributed interactions. When the two quantities agree, the system is said to be in thermal equilibrium.

Checking when a simulation is in equilibrium is an important part of transitioning from learning about the physical characteristics of a system to developing optimizers based on these finite-temperature simulation techniques. As shown in Fig. 3.5, if one can thermalize the system at a low enough temperature, one can reach a solution more quickly due to the higher probability of being in a ground state.

3.3 Optimization

Optimization is the process of minimizing (or maximizing) a mathematical function by choosing input values from a set and computing the function. The optimization of spin glasses provides a large test bed of problems for benchmarking algorithms and novel computing designs due to their NP-complete complexity. A number of heuristics, as well as exhaustive search methods, have been designed and developed to minimize spin-glass Hamiltonians as efficiently as possible. There exists exact methods such as "branch and cut" which will find the true optimum. However, these tend to be slow [64]. My focus is on heuristic methods, which includes simulated annealing, parallel tempering, popula-

tion annealing, genetic algorithms, and evolutionary algorithms [65, 66]. In the last two decades, quantum heuristics have been proposed as an alternative to classical heuristics, due to their potential to exploit quantum superposition and quantum tunneling effects. This section introduces the most generic optimization algorithm, simulated annealing, followed by an optimization method used by current quantum annealing devices, and finally parallel tempering as an optimizer.

3.3.1 Simulated Annealing

Simulated annealing (SA) is the most generic of optimization algorithms inspired by the cooling of a crystal to avoid defects. One stochastically samples the cost function $\mathcal{H}(\mathcal{S})$ at finite temperature via simple Monte Carlo to obtain a stationary state described by the Boltzmann distribution. Once the system is in thermal equilibrium, we reduce the temperature and equilibrate again [67, 68]. It was proven by Geman and Geman that infinitely slow cooling will find the true optimum of the cost function [69]. Simulated annealing is a sequential optimization algorithm because the temperature of the system can only decrease, unlike parallel tempering. If the energy landscape is rough, it is likely unable to escape metastable states. To overcome this problem, one can perform many rapid quenches of the temperature and repeat the sampling many times to try many paths through the energy landscape and find the true optimum. An example of simulated annealing pseudo-code is given in Alg. 1.

3.3.2 Quantum Annealing

Similar to simulated annealing, quantum annealing (QA) uses quantum tunneling and quantum fluctuations instead of thermal fluctuations to overcome energy barriers. Due to recent production of quantum annealing hardware, there is an incredible interest in the study of Ising spin-glasses [26, 30, 31, 70, 24, 25, 71, 27, 28, 29, 23, 32, 33, 34]. Instead of quenching the temperature, one quenches quantum fluctuations. Quantum annealing is

Algorithm 1 Simulated Annealing

begin
choose start configuration $\mathcal{S}$
for $t = 1, ..., t_{max}$ **do**
    begin
    set temperature $T = T(t)$
    Monte Carlo(N_{MCS}, T) at temperature T with N_{MCS} sweeps
    end
end

not limited to a local search via Monte Carlo. The strength of the fluctuations determines the length of tunneling in phase space. Tall barriers in the energy landscape, which can be difficult for thermal fluctuations to overcome, can now be tunneled through if their widths are small enough.

In quantum annealing, one modifies a classical Ising Hamiltonian by using Pauli spin matrices and adding a kinetic term for the quantum fluctuations;

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J_{ij} S_i^z S_j^z - D \sum_i S_i^x, \tag{3.13}$$

where J_{ij} is the interaction between two spins and D, the strength of quantum fluctuations. Note that the spin matrices do not commute, e.g., $[S_i^z, S_i^x] \neq 0$.

In order to minimize the Hamiltonian, one starts at zero temperature and a large value of D to randomize the configuration via quantum fluctuations alone. Then, one successively reduces the transverse field D via an annealing protocol, $D(t)$, until the system reaches the equilibrium state at $D = 0$. Again, if one reduces the quantum fluctuations infinitely slowly, the method will converge to the true optimum [70].

The latest D-Wave 2000Q quantum annealing device operates at a temperature of $1.5\mathrm{mK}$. This finite temperature might assist quantum annealers in finding optimal solutions through the use of both quantum and thermal fluctuations [72, 73]. We explore the

effects of finite-temperature fluctuations in Sec. 4.3. An example of quantum annealing pseudo-code is given in Alg. 2.

Algorithm 2 Quantum Annealing

begin
choose start configuration $\mathcal{S}$
for $t = 1, ..., t_{max}$ **do**
    set fluctuation strength $D = D(t)$
end
readout

3.3.3 Parallel Tempering as an Optimizer

Simple Monte Carlo combined with parallel tempering has, until recently, been the state-of-the-art algorithm for finding the ground-state configuration of an Ising spin-glass Hamiltonian [74, 75, 76]. For low enough temperatures, the ground state E_0 is the most probable state for the system. By implementing temperatures high enough to ensure proper mixing of configurations and low enough for simple Monte Carlo to likely dip into the ground state, one can create an optimization algorithm to find the minimum of a cost function. A general conclusion is that Monte Carlo heuristics based on thermal annealing are enhanced by mechanisms that improve thermalization at every temperature. In parallel tempering this mechanism is replica exchange [59, 77, 55].

Parallel tempering has proven to be a versatile optimizer in many research fields such as physics, chemistry, and biology [57]. An application of Parallel Tempering won the 2016 Max-SAT competition in several problem categories [38].

Parallel tempering is still a local update optimizer, meaning spins are flipped one at a time. In Ch. 4, we introduce a cluster algorithm that allows for large rearrangements of

Algorithm 3 Parallel Tempering

begin
choose start configuration $\mathcal{S}_1$ and $\mathcal{S}_2$
for $t = 1, ..., t_{max}$ **do**
  begin
  Monte Carlo(N_{MCS}, T_1) for system 1
  Monte Carlo(N_{MCS}, T_2) for system 2
  $\Delta E = \mathcal{H}(\mathcal{S}_2) - \mathcal{H}(\mathcal{S}_1)$
  if $\Delta E > 0$ **then**
    accept $[\mathcal{S}_1\, \mathcal{S}_2] \rightarrow [\mathcal{S}_2\, \mathcal{S}_1]$
  else
    begin
    $w = \exp(-(\beta_1 - \beta_2)\Delta E)$
    generate uniform random number $x \in [0, 1]$
    if $x < w$ **then**
      accept $[\mathcal{S}_1\, \mathcal{S}_2] \rightarrow [\mathcal{S}_2\, \mathcal{S}_1]$
    end
  end
end

spins to effectively "teleport" through tall energy barriers to improve parallel tempering. An example of parallel tempering pseudo-code is given in Alg. 3.

4. BENCHMARKING OF NOVEL CLASSICAL ALGORITHMS*

The best benchmark to test the performance of an optimization algorithm is a spin glass [78]. Both the disorder and frustration produce a complex energy landscape that is difficult for optimization algorithms. As such, all current benchmarks of quantum annealing machines attempt to find the ground state of an Ising spin glass. Due to diverging equilibration times in Monte Carlo simulations of spin glasses at low temperatures, it is important to use fast algorithms in order to probe the ground-state manifold. In this chapter, we introduce a novel algorithm for improving equilibration time through the use of Houdayer-like cluster moves that was extended from two-dimensions to any space dimension [60]. In addition, this algorithm also improves sampling of ground states in problems with degenerate solutions. Houdayer's cluster moves do not change the total energy of a system of two replicas. Thus, if both replicas are in the minimum-energy state are subject to a Houdayer cluster move, their energies can not change and two new solutions are generated. This method is exploited to produce a new algorithm that improves sampling of ground states from solutions produced by quantum annealing devices which are known to have biased solutions [2]. Finally, we introduce an algorithm to improve the recovery of ground-state solutions from simulations affected by analog-noise based on analytical results by Nishimori [41].

4.1 A Cluster Algorithm for Non-Planar Ising Spin Glasses

As explained in Ch. 3, spin systems with disorder and frustration are difficult to study both analytically and numerically. Simulations on ferromagnetic models with $J_{ij} = 1$ benefit greatly from cluster algorithms such as the Wolff cluster algorithm [58]. However, a cluster algorithm for generic spin-glass systems remained elusive. Presented here

*Part of this section is reprinted from Ref. [1]. Copyright 2015 by the American Physical Society.

*Part of this section is reprinted from Ref. [72]. Copyright 2016 by the American Physical Society.

is an algorithm based on the Houdayer cluster algorithm [60], from Sec. 3.1.2, for two-dimensional spin glasses, that leads to a speedup over conventional state-of-the-art methods that improves with system size.

The method updates large patches of spins at once, effectively randomizing the configurations and efficiently overcoming large barriers in the free-energy landscape. Furthermore, the energy of the system remains unchanged when performing a cluster move. This means that the numerical overhead is very small because the rejection rate is zero and there is no need to, for example, compute any random numbers for a cluster update. The use of cluster moves makes it possible to obtain a speedup of several orders of magnitude in two-dimensional systems, allowing one to simulate considerably larger system sizes.

In this section, we show that Houdayer-like cluster moves can be applied to spin systems on topologies where the percolation threshold is below 50% provided that the interplay of temperature and frustration prevents clusters from spanning the whole system by restricting the temperatures where cluster moves are performed. The inherent frustration present in the spin-glass Hamiltonian prevents clusters from spanning the whole system for temperatures below the characteristic energy scale of the problem.

4.1.1 Models and Observables

The Hamiltonian of a generic Ising spin glass without a field is defined by

$$\mathcal{H} = \sum_{\langle i,j \rangle} J_{ij} S_i S_j \qquad \text{with} \qquad S_i \in \pm 1. \tag{4.1}$$

The interactions J_{ij} are chosen from a normal Gaussian distribution shown in Eq. (2.3) with mean $\mu = 0$ and variance $\sigma^2 = 1$. In order to determine if thermalization is improved, we follow the recipe of Sec. 3.2 and record the average energy $[\langle E \rangle]$ and link overlap q_ℓ from Eq. (3.10). Recall that one can equate the internal energy per spin to the internal

energy computed from the link overlap [63].

$$E(q_\ell) = -\frac{\sigma^2}{T}\frac{N_b}{N}(1 - q_\ell) \tag{4.2}$$

where N is the number of spins, and N_b is the number of interactions between spins. Thus, to test that the system is thermalized we study the time dependent behavior of

$$\Delta = [\langle E(q_\ell)\rangle - \frac{\langle E\rangle}{N}] \tag{4.3}$$

When $\Delta \to 0$, the bulk of the disorder instances are thermalized [79]. Simulation parameters are listed in Table 4.1.

Table 4.1: Parameters of the simulation of isoenergetic cluster moves in two space dimensions (2D), three space dimensions (3D), and on the chimera (Ch) topology. For each topology simulated and system sizes N, we compute $N_{\rm sa}$ disorder instances and measure over 2^b Monte Carlo sweeps (and isoenergetic cluster moves) for each of the $2N_T$ replicas. $T_{\rm min}$ [$T_{\rm max}$] is the lowest [highest] temperature simulated, and N_T is the total number of temperatures used in the parallel tempering Monte Carlo method. Isoenergetic cluster moves only occur for the lowest $N_{\rm c}$ temperatures simulated (determined from Fig. 4.2).

	N	$N_{\rm sa}$	b	$T_{\rm min}$	$T_{\rm max}$	N_T	$N_{\rm c}$
2D	256, 576, 1024	10^4	22	0.2120	1.6325	30	30
Ch	128, 288, 512, 800, 1152	10^4	22	0.2120	1.6325	30	19
3D	64, 216, 512, 1000, 1728	1.5×10^4	23	0.4200	1.8000	26	13

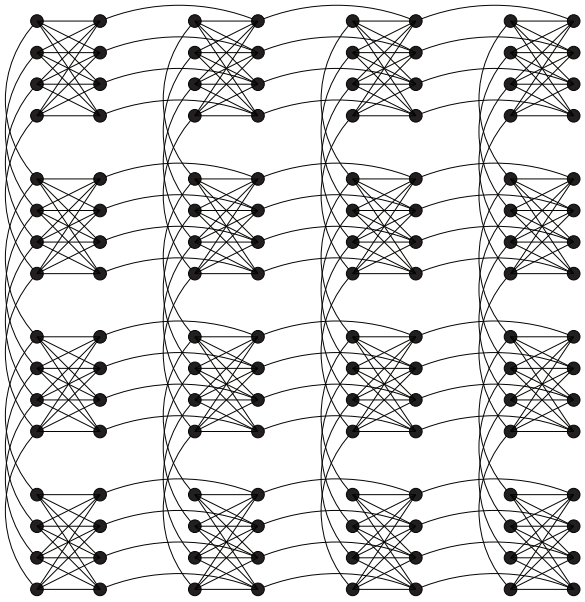

Figure 4.1: Chimera topology with $N = 128$ sites [10]. The circles denote spins and lines between circles denote interactions.

4.1.2 Algorithm

We introduce the *isoenergetic cluster algorithm* (ICA) for spin-glass Hamiltonians in any space dimension [1]. These rejection-free cluster moves accelerate thermalization by several orders of magnitude, even for systems with space dimensions larger than two dimensions such as the three-dimensional cubic lattice and the Chimera lattice shown in Fig. 4.1. The interplay of temperature and frustration prevents clusters from spanning the entire system despite having percolation thresholds below $p = 0.5$. In Fig. 4.2, we see that clusters can remain below the percolation threshold at approximately $T \sim J$, where

J is the energy scale of the associated problem. We can exploit this feature and restrict Houdayer cluster moves to the temperature region where they will be most efficient.

One simulation step using isoenergetic cluster moves follows:

- Perform one Monte Carlo sweep (N Metropolis updates) for each replica.
- If the cluster size is greater than half the spins, that all of the spins in that configuration are flipped thus reducing the cluster size while leaving the energy unchanged.
- Perform one Houdayer cluster move for all temperatures $T \leq J$.
- Perform one parallel tempering update for a pair of neighboring temperatures.

It is important to reiterate, the main difference lies in applying the cluster moves to *only* the temperatures where the isoenergetic cluster moves are efficient and reducing the cluster size to reduce numerical overhead.

Simulations on the chimera lattice show the overhead of ICA over PT is approximately 25% and is roughly independent of the system size for the studied N. However, the overhead for the HCA over PT is at least 50% and grows with increasing system size.

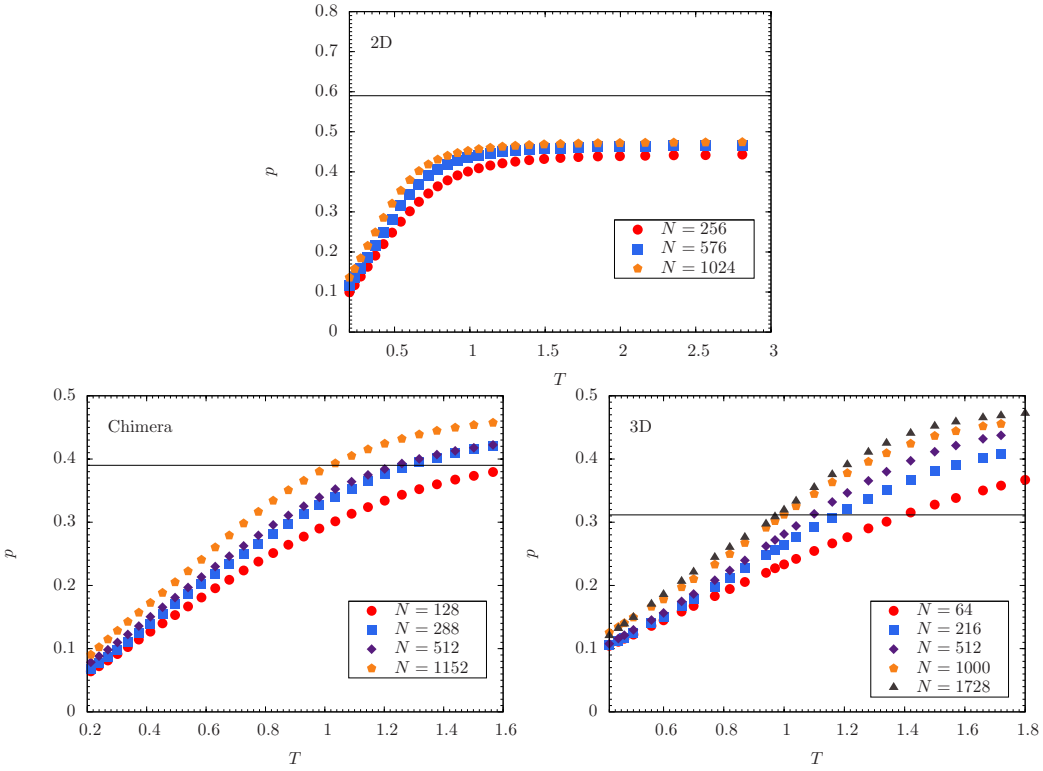

Figure 4.2: (Top panel) Fraction of spins p of potential cluster sites as a function of temperature T for different system sizes N in the two-dimensional square lattice. The horizontal line represents the percolation threshold of a two-dimensional square lattice, i.e., $p_c \approx 0.592$ [61]. Because $p \to 0.5$ for $T \to \infty$, for all T clusters do not percolate, which is why the ICA is efficient in two-dimensional planar geometries. (Center panel) p as a function of temperature T for different system sizes N on the chimera topology. The horizontal line represents the percolation threshold of the non-planar chimera topology shown in Fig. 4.1, namely $p_c \approx 0.387$ [80]. For $T \gtrsim J = 1$ clusters percolate and cluster updates provide no gain. (Bottom panel) p as a function of temperature T for different system sizes N in three space dimensions (3D). The horizontal line represents the percolation threshold of the three-dimensional cubic lattice ($p_c \approx 0.311$ [62]). For $T \gtrsim J = 1$ clusters percolate. In all panels, error bars are computed via a jackknife analysis over configurations and are smaller than the symbols.

4.1.3 Effects of Percolation on Cluster Moves

Figure 4.2 shows the fraction of potential cluster members as a function of temperature T for different system sizes with N spins and for three different topologies. The top panel

of Fig. 4.2 shows data for a two-dimensional square lattice where the percolation threshold is known to be $p_c \approx 0.592$ [61]. Thus, for all temperatures simulated, the fraction of cluster sites is below the percolation threshold and saturates at 50% for $T \to \infty$. This means that cluster updates are efficient for all temperatures because the clusters never percolate. One would expect the clusters to percolate in higher dimensions as connectivity is increased. However, in spin glasses this is not the case due to the frustration present as shown in the second and third panel in Fig 4.2. For the chimera topology and in the three-dimensional cubic lattice, for increasing system size the fraction of cluster sites converges to a limiting curve that crosses the percolation threshold at approximately $T \approx J = 1$. This means that, for all $T \geq J$, clusters percolate and cluster updates are simply numerical overhead. However, for $T \leq J$, the fraction of cluster sites is below the percolation threshold and cluster moves in this temperature regime should improve thermalization.

When the interactions J_{ij} are drawn from a normal Gaussian distribution, the ground state is unique. There is only one configuration $\mathcal{S}$ that minimizes the Hamiltonian. In Fig. 4.2 the fraction of spins potentially in a cluster also approaches zero for $T \to 0$. Therefore, a cluster is composed of no sites, or the entire lattice. In the case of disorder distributions that yield highly degenerate ground states, such as the bimodal distribution in Eq. (2.2), it is possible to continue to have clusters at zero temperature. Thus, it is possible to hop around the ground state manifold by only applying cluster moves at zero temperature. An interesting application of this feature is to "fix" the poor sampling of the D-Wave device by repeatedly performing cluster moves on the configuration results to produce new solutions shown in 4.2.5.

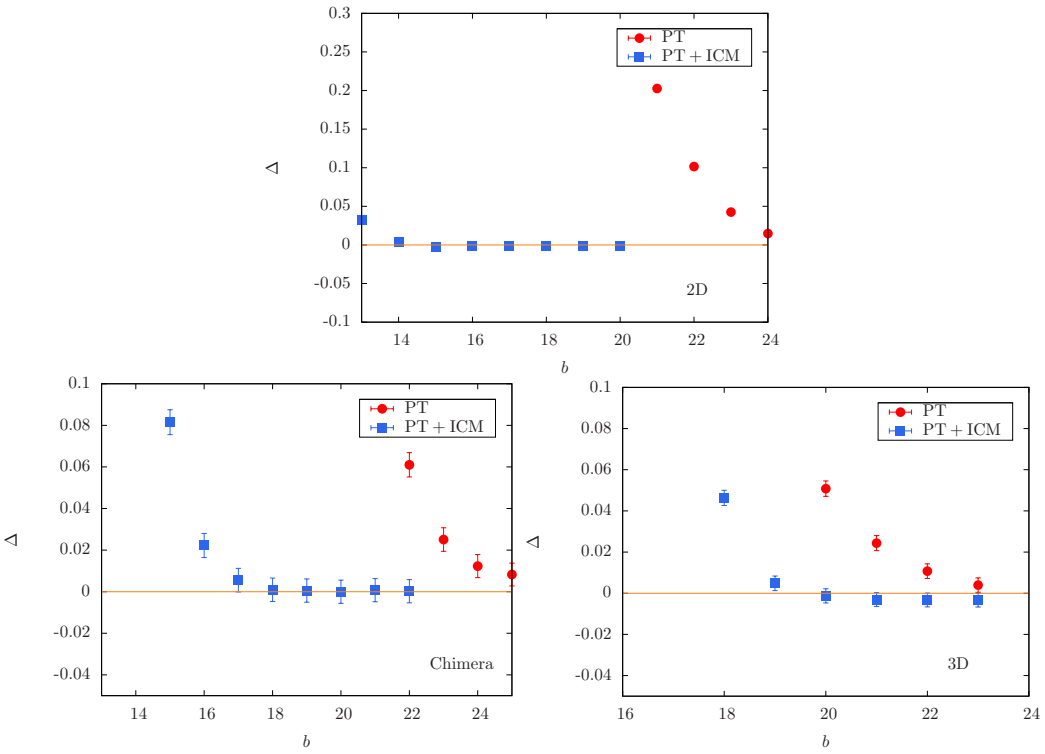

Figure 4.3: (Top panel) Δ [Eq. (4.3)] as a function of simulation time $t = 2^b$ measured in Monte Carlo sweeps in two space dimensions (2D) for $N = 1024$ and $T = 0.212$. Simulations using vanilla PT thermalize at at least 2^{25} Monte Carlo sweeps, whereas with the addition of ICA, thermalization is reduced to approximately 2^{16} Monte Carlo sweeps. This means approximately 2 orders of magnitude improvement. (Center panel) Δ as a function of simulation time $t = 2^b$ measured in Monte Carlo sweeps for an Ising spin glass on chimera with $N = 1152$ spins at $T = 0.212$. Simulations using PT thermalize at approximately 2^{25} Monte Carlo sweeps, whereas the addition of ICA reduces thermalization to 2^{18} Monte Carlo sweeps. Again, approximately 2 orders of magnitude speedup. (Bottom panel) Δ as a function of simulation time $t = 2^b$ measured in Monte Carlo sweeps in three space dimensions (3D) for $N = 1728$ and $T = 0.42 \sim 0.43 T_c$. Using standard PT, the system thermalizes approximately after 2^{23} Monte Carlo sweeps. This time is reduced to $\sim 2^{20}$ Monte Carlo sweeps when ICA is added. In all panels, error bars are computed via a jackknife analysis over configurations.

4.1.4 Results

We compare the thermalization time of parallel tempering Monte Carlo with the Isoenergetic Cluster Algorithm at the lowest temperature of the simulation where, in principle,

thermalization is most difficult.

Figure 4.3 shows Δ as a function of Monte Carlo time measured in lattice sweeps $t = 2^b$. The first panel shows data in two space dimensions for simulations using ICA and simple parallel tempering Monte Carlo for $N = 1024$ spins at $T = 0.212$. Clearly the inclusion of cluster moves show an improved thermalization as can be expected from the simpler Houdayer's cluster algorithm. The second panel shows data on the chimera topology with $N = 1152$ spins and $T = 0.212$, where the HCA is not expected to show any improvement due to the percolation threshold $p_c < 0.5$. ICA clearly improves thermalization in comparison to PT by at least two orders of magnitude, an amount that increases with system size. Finally, the last panel shows Δ as a function of simulation time in three space dimensions with $N = 1728$ spins and $T = 0.42 \ll T_c$. The data show a speedup of approximately one order of magnitude that again grows with increasing system size.

Figure 4.4 shows the ratio of thermalization time using PT and using PT with ICA for different topologies at the lowest simulation temperature as a function of system size. In all cases, the speedup increases with increasing system size.

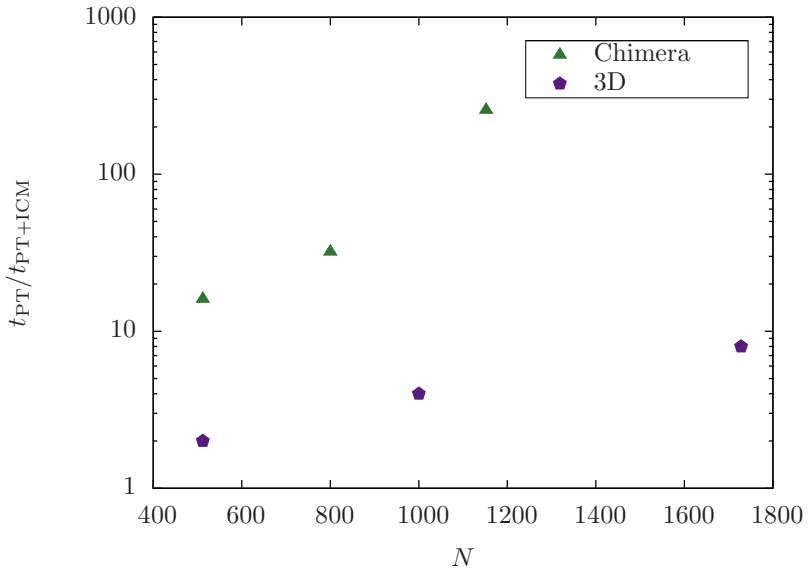

Figure 4.4: Ratio between the *approximate* average thermalization time of PT and PT+ICM for different topologies at the lowest simulation temperature (see Table 4.1) as a function of system size N. In all cases the speedup increases with increasing system size. Note that thermalization times have been determined by eye.

4.1.5 Summary and Applications

I have presented a rejection free cluster algorithm for spin glasses in any space dimension that greatly improves thermalization. By restricting Houdayer cluster moves to temperatures where cluster percolation is hampered by the interplay of frustration and temperature, we are able to extend the Houdayer cluster algorithm for two-dimensional spin glasses to any topology or space dimension. This implementation of the cluster updates represents only a minor overhead compared to the thermalization time speedup, that improves with system size, obtained from the isonenergetic cluster algorithm. This algorithm represents a new state-of-the-art in the benchmarking of quantum annealing devices [37]. Isoenergetic cluster moves have been used in studies of new benchmark problems [19, 37], the basis of a new algorithm for satisfiability problems [38], ground state sampling studies [36, 2], and studies of the effects of noise on analog quantum devices [4].

4.2 Improving Fair Sampling in Non-Planar Topologies

The thermalization of a spin glass represents a first stringent metric when evaluating the quality of a novel optimization approach. However, verifying that an optimizer can sample all solutions that minimize the Hamiltonian is a far more stringent test for any newly developed algorithm. Many different algorithms have been proposed to solve this optimization problem is quantum annealing [26], simulated annealing [67], parallel tempering [74, 75], and population annealing [81, 76]. In addition, uniform sampling of ground states is imperative for other combinatorial optimization problems in computer science such as propositional model counting (#SAT) [82], knapsack solution problem counting [83], and satisfiability-based set membership filters [84].

In previous work by Moreno *et al.*, parallel tempering was shown to be more efficient than simulated annealing at finding spin glass ground-state configurations with near equal probability [74]. In more recent work by Matsuda *et al.*, simulated quantum annealing was shown to perform worse than simulated annealing with certain ground state solutions being exponentially suppressed [9]. In addition, work by Wang *et al.* shows that population annealing is comparable to parallel tempering [85]. This theory is later experimentally demonstrated on the D-Wave 2X quantum annealer [86]. More complex driving Hamiltonians, which introduce quantum transitions between all states with equal weights, are proposed for future quantum annealing machines to ensure a fair sampling of the ground-states.

The newly developed ICA, explained in detail in the previous section, which combines parallel tempering with isoenergetic cluster moves allows for a wide-spread sampling of search space. Here, ICA is shown to improve the equal sampling of ground state configurations for the Chimera topology.

4.2.1 Model

We begin with an Ising spin glass model on the non-planar Chimera graph shown in Fig. 4.1. Its non-planar topology makes finding ground states of the Ising spin glass defined on a Chimera graph a worst-case NP-hard problem. The Hamiltonian for this spin glass model is given by

$$\mathcal{H} = \sum_{\langle i,j \rangle} J_{ij} S_i S_j \qquad \text{with} \qquad S_i \in \pm 1. \tag{4.4}$$

The coupling constants, $J_{ij} \in \{\pm 1, \pm 2, \pm 4\}$ are selected based on the range of ground-state degeneracy our high-performance computing cluster can store during a simulation. For example, the simplest choice of $J_{ij} \in \{\pm 1\}$ has ground state degeneracies of the order of 10^6 to 10^9 which requires a significant amount of simulation memory. Adding the additional values of ± 2 and ± 4 reduces the degeneracy to something tractable on the order of 10^2 to 10^5.

4.2.2 Metric for Improvement

To illustrate improvement of sampling we must first define a metric for evaluating the quality of sampling. Suppose n is the total number of times that ground states are found for an instance with ground-state degeneracy G. The probability distribution for finding any particular ground-state configuration is a binomial distribution. In the case of theoretically perfect sampling, if $p = 1/G$ and $q = 1 - p$ is the probability of failure in a binomial trial, then the expected number of successes in n trials is $e = np$ and the variance of the binomial distribution is $\sigma^2 = npq$. Thus, the theoretical relative standard deviation Q_{th}, is

$$Q_{\text{th}} = \sigma/e = \sqrt{(1-p)/np} = \sqrt{(G-1)/n}. \tag{4.5}$$

The algorithm is optimal if the numerical relative standard deviation of the frequency of ground state configurations, Q_{num}, is close or equal to the theoretical value, or if $Q_{\mathrm{num}}/Q_{\mathrm{th}} = 1$. In practice, Q_{num} for any algorithm is greater than the theoretical value Q_{th} due to finite computational resources.

Table 4.2: Parameters of the simulation for fair sampling of ground states using isoenergetic cluster moves. For each instance class and system size N, we compute N_{sa} instances. $N_{\mathrm{sw}} = 2^b$ is the total number of Monte Carlo sweeps for each of the $4N_T$ replicas for a single instance, T_{min} [T_{max}] is the lowest [highest] temperature simulated, and N_T and N_{hc} are the number of temperatures used in the parallel tempering method and in the isoenergetic cluster algorithm, respectively.

Topology	N	N_{sa}	b	T_{min}	T_{max}	N_T	N_{hc}
2D	144	360	24	0.0500	3.0500	35	35
2D	256	360	24	0.0500	3.0500	35	35
2D	576	322	24	0.0500	3.0500	35	35
2D	784	232	24	0.0500	3.0500	35	35
2D	1024	370	24	0.0500	3.0500	35	35
Chimera	128	360	24	0.0500	3.0500	35	20
Chimera	288	360	24	0.0500	3.0500	35	20
Chimera	512	360	24	0.0500	3.0500	35	20
Chimera	800	360	24	0.0500	3.0500	35	20
Chimera	1152	223	24	0.0500	3.0500	35	20

4.2.3 Isoenergetic Cluster Algorithm as a Solver

To determine if we have found the ground state solution of a system, we randomly initialize four replicas with the same disorder and keep track of the lowest energies of each

individual replica at the lowest temperature. If after the first half of the simulation, $N_{sw}/2$, the lowest found energies of the replicas are equivalent, we are confident the ground-state energy has been found and we record the frequency of ground states for the remaining half of the simulation. We choose to make sure that each configuration is recorded a minimum number of 50 times in order to increase our confidence that all accessible ground states have been found. The simulation parameters are shown in Table 4.2.

4.2.4 Results

Figure 4.5 shows $Q_{\rm num}\sqrt{n}$ as a function of ground-state degeneracy $G-1$ for different spin-glass instances on a chimera graph. The algorithm is claimed to be optimal if the data points from the numerical relative standard deviation are close to the theoretical line. It is clear that the data for ICA are closer to the theoretical dotted line than the data points from PT and this discrepancy increases as the system size increases.

In Fig. 4.6 we plot the median ratio of $Q_{\rm num}/Q_{\rm th}$ over different instances as a function of the system size N. $Q_{\rm num}/Q_{\rm th} \to 1$ implies optimal sampling. The data show that ICA performs better than PT and the improvement is more significant with increasing system size. Large median ratios $Q_{\rm num}/Q_{\rm th}$ for smaller system sizes are due to the choice of temperature set which has been specifically optimized for $N = 1152$. Note that statistical error bars are determined by a bootstrap analysis.

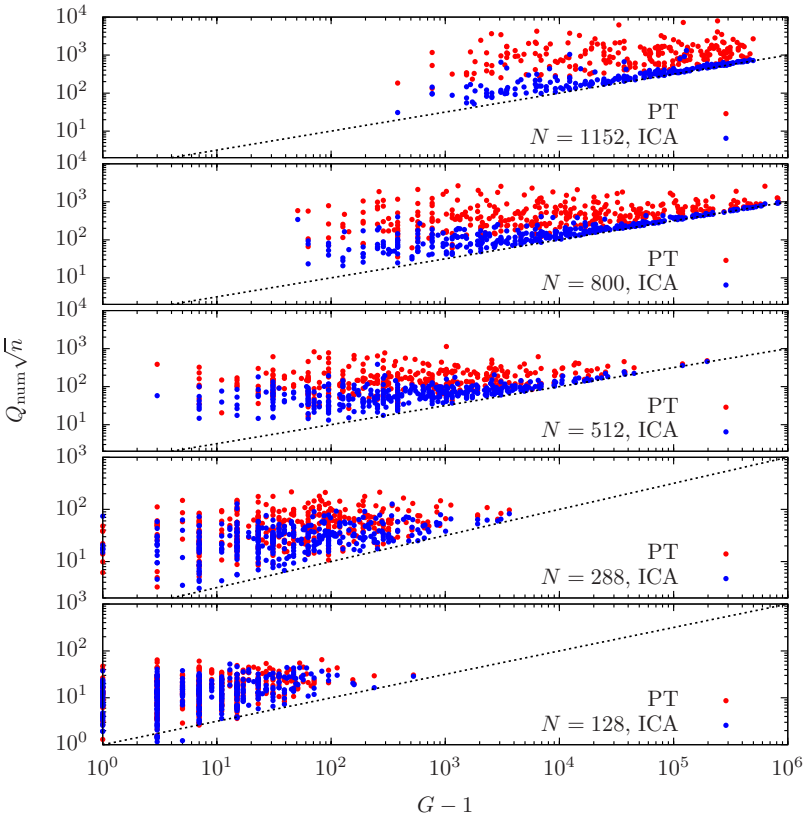

Figure 4.5: Scatter plot of quantities $Q_{\mathrm{num}}\sqrt{n}$ as a function of the ground-state degeneracy $G - 1$ for different spin glass instances with different system sizes N on a Chimera graph. The data points for ICA (blue color) are closer to the theoretical limit than those for the PT (red color), and this improvement gets better as the system size increases. The dotted line represents ideal uniform sampling of ground-state configurations, i.e., $Q_{\mathrm{num}}/Q_{\mathrm{th}} = 1$.

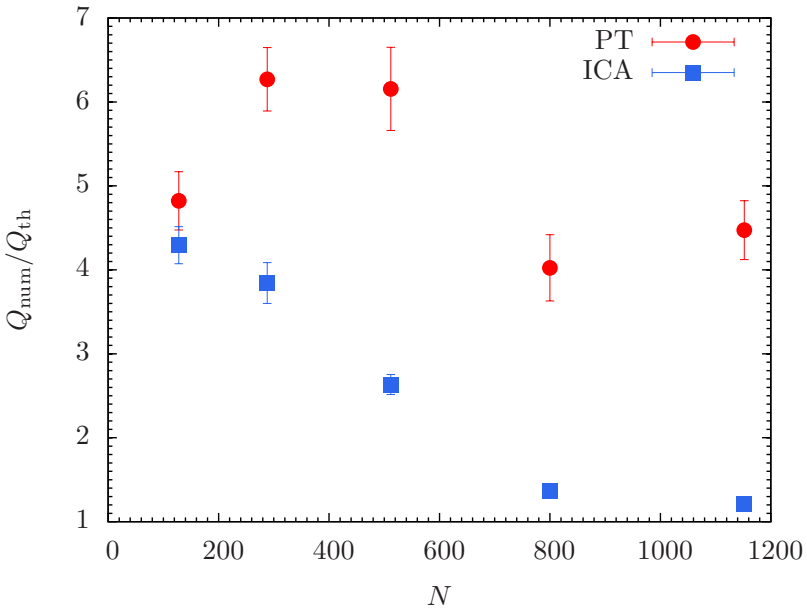

Figure 4.6: Median ratio $Q_{\rm num}/Q_{\rm th}$ over different spin glass instances as a function of the system size N on a Chimera graph. The data points show that ICA (blue color) performs better than PT (red color) for all system sizes and the gain is more significant with increasing system size. Note that the statistical error bars are determined by a bootstrap analysis.

In addition to the system size, we also investigate how the quality of fair sampling is related to ground-state degeneracy and plot $Q_{\rm num}/Q_{\rm th}$ as a function of ground-state degeneracy with the same system size $N = 800$. Figure 4.7 suggests that with more ground-state configurations, it is easier to sample those configurations with near-equal probabilities. Furthermore, careful examination of instances with the same system size and ground-state degeneracy suggests that $Q_{\rm num}/Q_{\rm th}$ is closely related to the Hamming distances between ground state configurations.

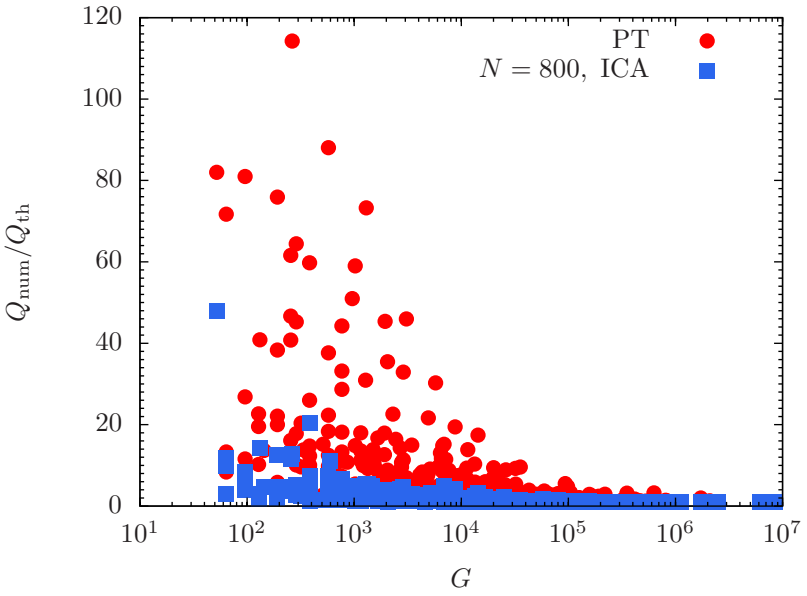

Figure 4.7: Scatter plot of ratio $Q_{\rm num}/Q_{\rm th}$ as a function of ground-state degeneracy G for different spin glass instances with system size $N = 800$ on Chimera graph. Both data for PT and ICA suggest that the more ground-state configurations, the easier to sample all ground-state configurations with near-equal probabilities.

Figure 4.8 shows that instances with large Hamming distances between the ground-state configurations have a higher $Q_{\rm num}/Q_{\rm th}$ due to having to flip a large number of spins in order to completely explore the ground state manifold. To visualize this, Fig. 4.9 shows two examples of ground-state configurations with different hamming distances on a Chimera graph with $N = 128$. In the ball-like ground-state manifold, exploration is easy with simple Monte Carlo because all the ground-state configurations are related by single spin flips. The square-like ground-state manifold benefits from the large rearrangements of spins provided by ICA to efficiently hop between the clusters of configurations related by single spin flips.

As the system size increases, ICA improves the sampling by having $Q_{\rm num}/Q_{\rm th}$ approach 1. This is a great improvement over quantum annealing, in which some solutions are exponentially suppressed, in other words, would require an exponential amount of time

to find that particular suppressed solution [9]. This is an important issue for the design of quantum annealing architectures.

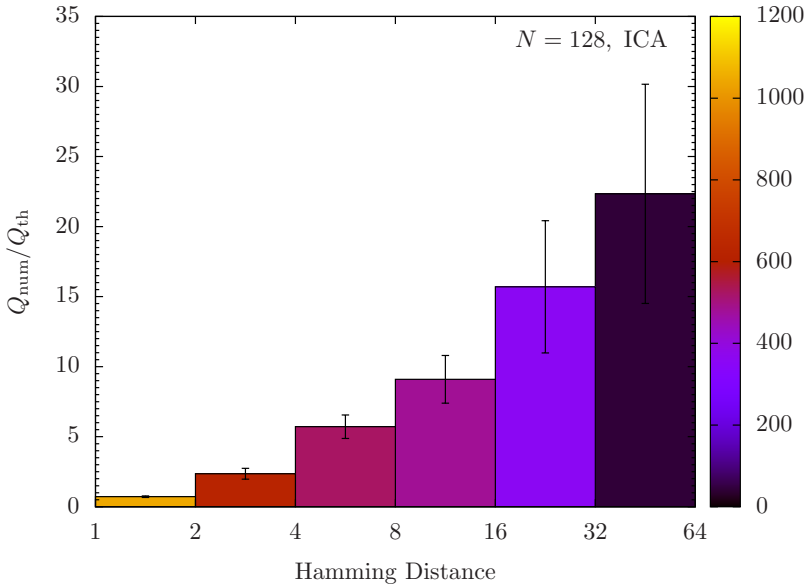

Figure 4.8: Median ratio $Q_{\rm num}/Q_{\rm th}$ as a function of Hamming distance for different spin glass instances with system size $N = 128$ and degeneracy $G = 4$ on Chimera graph. Data from ICA suggest that the smaller Hamming distance between ground-state configurations, the easier to sample all ground-state configurations with near-equal probabilities. Note that bar chart represents median ratios $Q_{\rm num}/Q_{\rm th}$ between Hamming distance 1-2, 2-4, 4-8, 8-16, 16-32 and 32-64, respectively. The statistical error bars are determined by a bootstrap analysis.

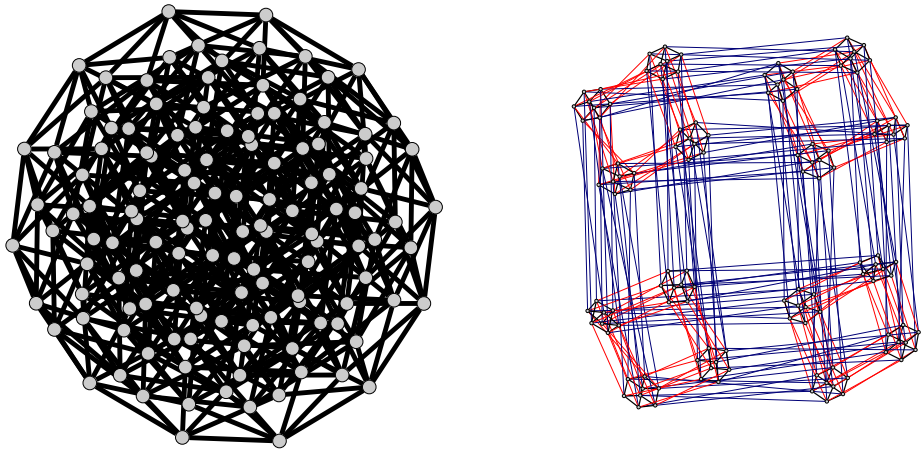

Figure 4.9: Two examples of ground-state configurations with different Hamming distances on a Chimera graph for system size $N = 128$. The Hamming distance denotes the difference between two binary strings (ground-state configurations). Each dot in the figure represents a ground-state configuration, black lines are 1-bit differences, red lines are 2-bit differences, and anything that is a light color or blue is an even greater difference. In the example on the top, all ground-state configurations are related by 1-bit differences, while the example on the bottom shows that Hamming distances between certain ground-state configurations can be large—which means that it will take longer for the system to move from one ground-state configuration to another and this will cause larger fluctuations in the ground-state frequency.

4.2.5 Generating New Solutions from Known Results

An alternative algorithm to improve the sampling of biased optimization schemes is to begin with a bank of known solutions, such as those found by quantum annealing. One can then perform only cluster moves without simple Monte Carlo or parallel tempering to potentially generate new ground-states at very little computational cost. Due to the change in energy of the Houdayer cluster moves $\Delta E = 0$ this means that $\Delta E_2 = -\Delta E_1$. However, because both configurations are already minimum energy solutions, this implies $\Delta E_1 = 0$, thus $\Delta E_2 = 0$ and the two new states are also minimum energy solutions. Due to the lack of Monte Carlo and parallel tempering, it should be noted, this method

is non-ergodic, meaning it will not find all the potential ground state configurations. The algorithm will simply generate more solutions if there are more to be found based on what is contained in the initial bank of solutions.

A description of the algorithm follows:

- Randomly choose two configurations from the bank of known solutions and compute the q-space configuration.
- Identify all clusters with $q_i = -1$.
- Flip a cluster in their original configurations to potentially generate two new solutions and add them to the bank.
- Repeat for the remaining identified clusters.

Initial configurations are chosen randomly because for some highly degenerate disorder distributions, the number of degenerate ground-state configurations, G, can be larger than 10^6 as in the case of the $J_{ij} \in \{\pm 1\}$ bimodal disorder.

4.2.6 Results

We apply the algorithm to a class of problems with well controlled degeneracy used to study exponentially-biased sampling of quantum annealing devices, in this case, the D-Wave 2X quantum annealer [2]. Interactions of this instance class, $J_{ij} \in \{\pm 5, \pm 6, \pm 7\}$, are chosen to limit the ground state degeneracy. Due to imperfections of the physical chimera graph and choice of interactions, instances typically have degeneracy of $G = 3 \times 2^k$ and those that fall outside this sequence are discarded. Table 4.3 gives the number of instances for each system size and total number of ground states of the simulation.

Table 4.3: Number of disorder instances from Mandrà *et. al* sorted by system size and number of ground states [2].

N	$G = 24$	$G = 48$	$G = 96$	$G = 192$	$G = 384$	$G = 768$	Total
512	63	51	48	26	0	0	188
648	70	56	59	75	0	0	260
800	28	52	32	59	38	6	215
968	22	15	31	30	28	21	147

Figure 4.10 shows results for different chimera subgraphs of the full $N = 1152$ chimera graph. The number of initial configurations N_{in} given by the results of the D-Wave simulation are read into the algorithm and pairs of solutions are randomly chosen in order to discover new solutions. For some instances, as few as $N_{\text{in}} = 10$ solutions are input and more than 100 solutions, N_{out}, are output. Points to the left of the $N_{\text{out}} = N_{\text{in}}$ line show significant improvement to the number of solutions produced by the D-Wave quantum annealer. In many cases for smaller system sizes, the algorithm can find the remaining solutions to an instance not found by the D-Wave quantum annealer.

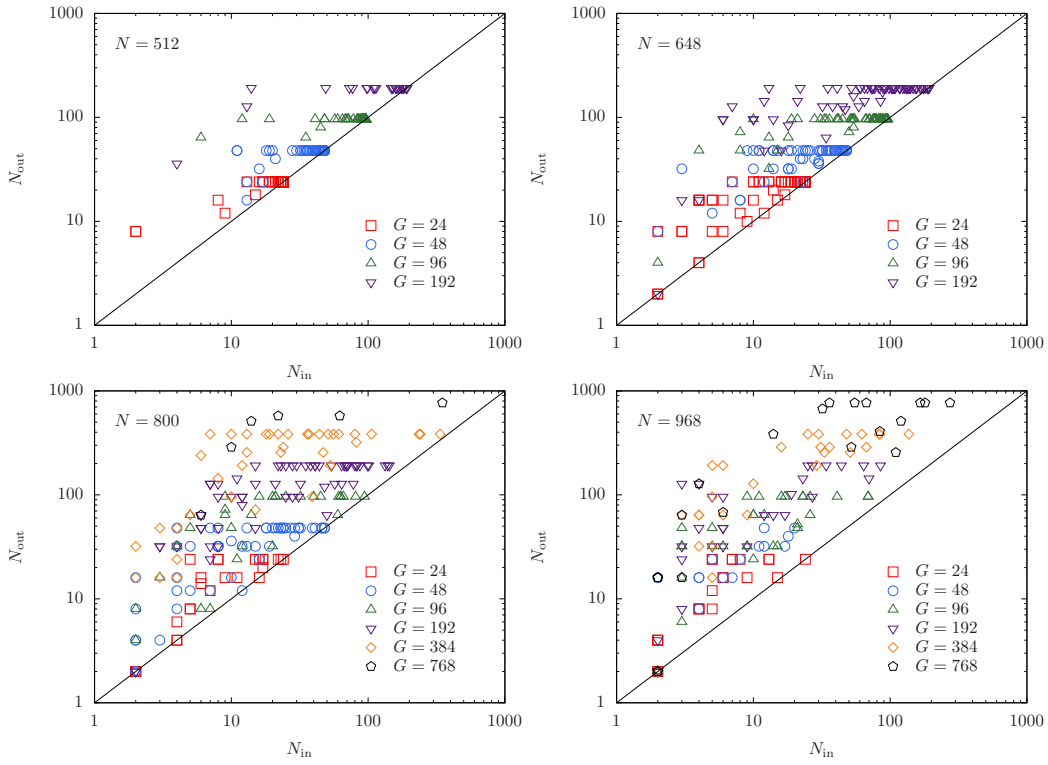

Figure 4.10: The number of solutions given by using only cluster moves, N_{out}, as a function of solutions given by the results of a D-Wave 2X quantum annealer simulation, N_{in}. Each point represents an individual instance with different disorder interactions. Points to the left of the $N_{\text{out}} = N_{\text{in}}$ solid line imply new solutions are found.

Figure 4.11 shows the ratio of improvement of instances sorted by total number of ground states G. The improvement increases with the total number of ground states and with system size. This is remarkable because in the original D-Wave quantum annealing simulation, the D-Wave produced 10^5 readouts. In other words, out of the 10^5 configurations output by the D-Wave only a few readouts were solutions and these solutions were biased to even fewer ground-state configurations. The ability of this algorithm to recover the remaining solutions will be useful in future studies requiring degenerate solutions.

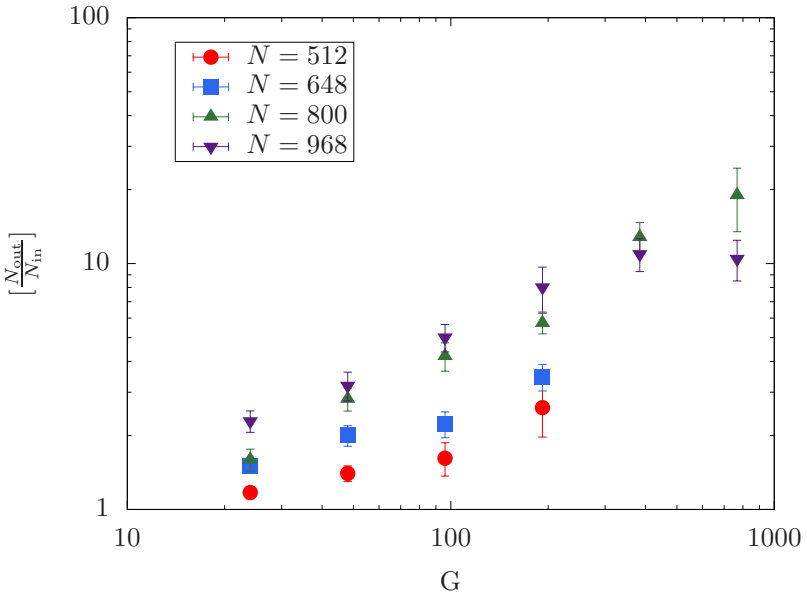

Figure 4.11: The average ratio of improvement of instances by number of ground states G for different chimera subgraphs. The improvement increases for the number of ground states and the system size. The crossing of $N = 800$ and $N = 968$ is likely due to the limited numbers of instances for these numbers of ground states.

Finding multiple solutions to a problem has many benefits in model counting [87] and SAT filter development [84, 88]. Multiple solutions may help hidden properties or relations of the problem being studied. For some real-world applications, there may be factors that are difficult to model mathematically. Additional solutions will allow decision makers to have more options for consideration of similar quality with factors that are not captured in simulations.

4.2.7 Summary

I presented two novel cluster algorithms for sampling ground-state configurations of spin glasses on a chimera graph as well as on a two-dimensional square lattice. Parallel tempering updates combined with rejection-free isoenergetic cluster moves create a robust ensemble that is able to sample both low and high-energy configurations and al-

low global moves through the rough energy landscape. A more equiprobable sampling of ground-state configurations has been achieved, which ensures all ground-state configurations for benchmarks can be found. We also found that degeneracy and Hamming distances between different ground-state configurations are closely related to the relative standard deviation of frequency with which the ground states are found. Ground states with large degeneracy and small Hamming distances have a lower relative standard deviation of frequency. In addition, I also present an algorithm for traversing the ground-state manifold to generate new solutions from a bank of known solutions. The average ratio of improvement increases with the total number of ground states of the system and the system size. This will be useful when algorithms known to be biased, such as quantum annealing, produce solutions and other solutions can be potentially deduced with minimal computational effort.

4.3 An Algorithm for Finite-Temperature Decoding

Current quantum annealing devices operate at a low finite temperature, about 15 millikelvin in the case of the D-Wave 2000Q, to facilitate coherence between the quantum flux qubits that act as spins in the quantum Ising Hamiltonian. Thermal fluctuations play an important role in the simulation of spin glasses and it is not immediately clear if the finite temperature of a quantum annealer aids in finding the ground state solution by exploiting thermal fluctuations to escape local minima in the energy landscape or if the finite temperature is more likely to remove a system from the ground state. In Ref. [89], Pál Ruján first examined the correspondence between error-correcting convolution codes and gauge invariant spin-glass models to show that the optimal way to recover an original message is by decoding the message at a finite temperature $T_N(q) > 0$, where q is the strength of the channel noise, and T_N is the Nishimori temperature.

The Nishimori temperature is the temperature at which for a given amount of disorder in a problem, average values of certain observables such as the internal energy of a system may be computed exactly. It comes from the study of gauge theory on finite-dimensional spin glasses which uses the symmetry of the system to derive a number of rigorous and exact results. For the case of Gaussian disorder, the Nishimori temperature is defined as σ^2/μ, where μ is the mean and σ^2 is the variance of the distribution.

Suppose the information to be sent is a configuration of Ising spins $\{S_i = \pm 1\}_{i=1,\ldots,N}$. A simple way to send the information is to send the encoded interactions, $J_{ij} = S_i S_j$. The receiver receives noisy interactions and then finds the ground state of the disordered Ising model Hamiltonian. It is important to note that even if a small percentage q of the interactions are incorrect, it is still possible to retrieve the true configuration as long as errors are isolated from each other because isolated frustration does not change the ground-state configuration. Ruján's result was rigorously proven by Nishimori via gauge transformation [41]. More recently the idea was extended to the Chimera graph with ferromagnetic interactions [90].

Preliminary results from K. Nishimura using transfer-matrix methods on the triangular ladder model suggest that the correct ground state configuration is optimally retrieved at the Nishimori temperature for simple systems [72].

4.3.1 Model

Here, we introduce a Hamiltonian with Gaussian distributed interactions and added bond noise ζ also chosen from a Gaussian distribution,

$$\mathcal{H} = -\sum_{\langle i,j \rangle} (J_{ij} + \zeta_{ij}) S_i S_j \qquad \text{with} \qquad S_i \in \{\pm 1\}. \tag{4.6}$$

where

$$P(J_{ij}) \propto \exp\left(-\frac{(J_{ij}-\mu)^2}{2\sigma^2}\right) \qquad \text{and} \qquad P(\zeta_{ij}) \propto \exp\left(-\frac{\zeta_{ij}^2}{2\gamma^2}\right). \tag{4.7}$$

The Nishimori temperature for this model with Gaussian disorder and additional Gaussian noise is the sum of the two variances, $\sigma^2+\gamma^2$, divided by the sum of the two means, $\mu+0$,

$$T_N = \frac{\sigma^2+\gamma^2}{\mu}. \tag{4.8}$$

Figure 4.12 shows the phase diagram of the two-dimensional Ising model with Gaussian disorder. The Nishimori line, from which the Nishimori temperature N_T can be calculated is shown in red [91]. At this temperature, it is more likely to improve the inference of solution to a simulation with added noise than at higher temperatures where due to thermal fluctuations, averages of observables $\langle \mathcal{O} \rangle \to 0$, or too low temperatures where spins are frozen in incorrect orientations.

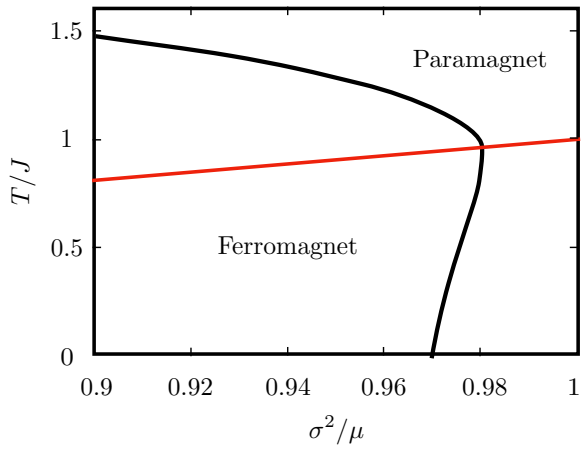

Figure 4.12: Illustration of the phase diagram of the two-dimensional random-bond Ising model with Gaussian disorder. The red line is the Nishimori line, from which one can determine the temperature given some amount of disorder. The boundary shown between these two regions is a guide to the eye. Adapted from Ref. [92].

4.3.2 Algorithm

To measure improved decoding, we measure the overlap of the sign of the average value of each spin of the noisy Hamiltonian with the ground state configuration $S_i^{(0)}$ of the noise-less Hamiltonian ($\zeta_{ij} = 0$).

$$M(T) = [S_i^{(0)} \cdot \mathrm{sign}\langle S_i^{(\gamma)} \rangle_T]_{\sigma,\gamma} \tag{4.9}$$

$M/N \to 1$ for $N \to \infty$ implies perfect decoding.

When $\sigma = 0$ the Hamiltonian represents the ferromagnetic Ising model. Then $S_i^{(0)} = 1$ and the overlap $M(T)$ is identical to that of the previous analytical results of Nishimori [41]. It is known in this case that the overlap takes a maximum value at $T_N = \gamma^2$. It is difficult to apply the same theory to the case where $\sigma \neq 0$ due to the lack of proper symmetry. The ground-state configuration is no longer uniform.

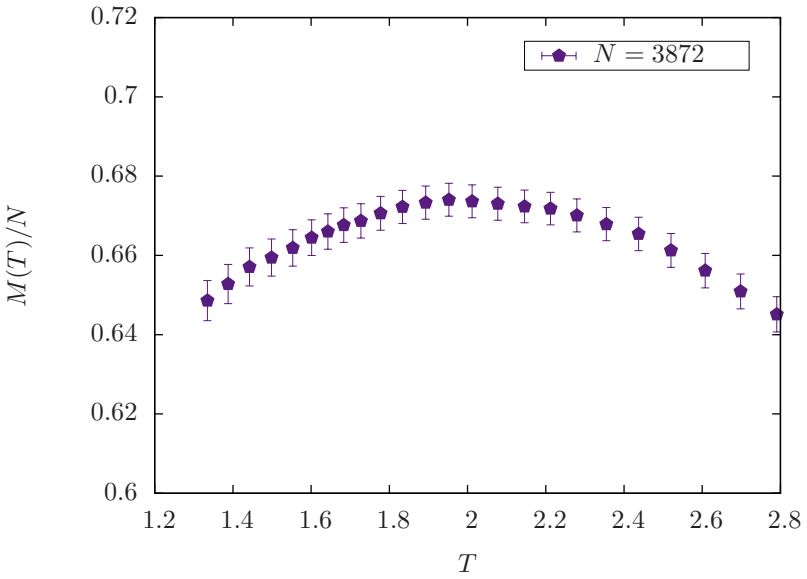

Figure 4.13: Results of finite-temperature decoding for the $N = 3872$ Chimera graph with $\sigma = 0$ and $\gamma = 1.4$. The data show a clear peak at the Nishimori temperature, $N_T = 1.96$ signifying improved decoding.

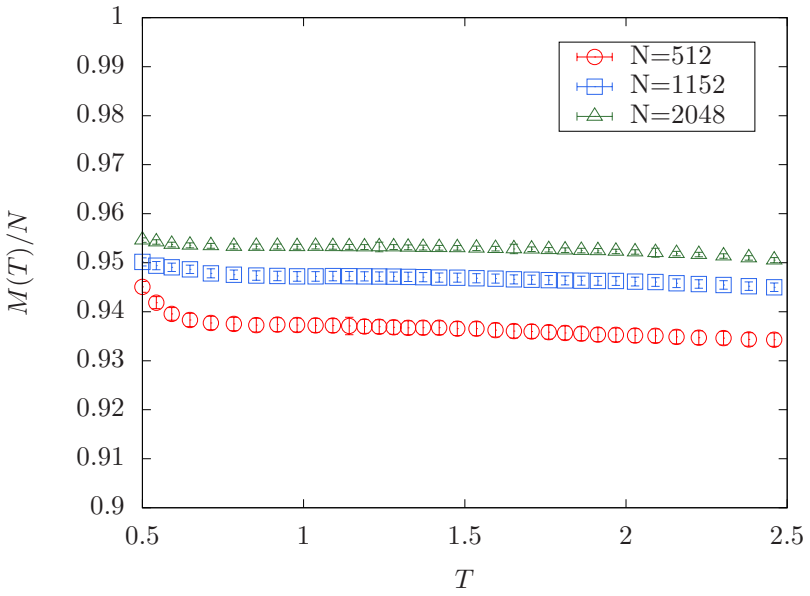

Figure 4.14: Null results of finite-temperature decoding for sizes of chimera graph with $\mu = 1$,$\sigma = 1$, and $\gamma = 0.5$. The data do not show improvement at the Nishimori temperature, $N_T = 1.25$.

4.3.3 Results

Preliminary results shown in Fig. 4.13 for the Chimera topology with $N = 3872$ spins with $\sigma = 0$ and $\gamma = 1.4$ show a clear improvement in decoding at the Nishimori temperature. This is in agreement with Nishimori's analytic results for improved decoding. However, when $\sigma \neq 0$ as in Fig. 4.14, this observable is no longer useful when using Monte Carlo simulations due to spin reversal symmetries and optimization methods that use multiple replicas and random initial configurations. Attempts at finding a metric in this regard were unfruitful. A brief description of the pitfalls of this metric for non-trivial ground states follows.

We begin by considering the simple ferromagnetic Ising model. Typical simulations using simple Monte Carlo with parallel tempering begin by initializing the replicas at each temperature with random configuration states. When the simulation is allowed to equili-

brate, one finds that due to each replica's path through phase space, each replica can be in either the all $S_i = 1$ or all $S_i = -1$ configuration that both minimize the Hamiltonian. One should take care when measuring $\langle S_i \rangle$ when using parallel tempering because neighboring temperatures with opposite configurations will cause $\langle S_i \rangle \to 0$. To fix this issue, in the case of the ferromagnetic Ising model, one can measure the magnetization of a replica and make a choice of ground state solution before recording S_i as a method of determining whether to record S_i or the opposite, $-S_i$, according to the sign of the magnetization as shown in Fig. 4.13.

Unfortunately, this fix no longer applies to the case when the ground state is not uniform. Spin-glass states are indistinguishable from random configurations, thus any attempts to record S_i or the opposite $-S_i$ result in biasing all spins and yield $M(T)/N = 1$ as $T \to \infty$ which is an absurdity. An alternative method to fix this problem is to calculate the overlap with the known ground state and "flip" the configuration accordingly depending on the sign of the overlap. However, this method does not yield any signs of improved decoding as shown in Fig. 4.14.

4.3.4 Summary

Finite temperature decoding presents an opportunity to use thermal fluctuations to help improve the ability to infer better solutions in problems affected by noise. I presented results for a trivial ferromagnetic distribution with Gaussian noise that agrees with analytical results. However, a metric for improved decoding for more complicated distributions remains to be discovered. Due to the non-trivial nature of spin-glass benchmarks, the current metric is not useful. This idea has recently been extended to finite quantum fluctuations in the quantum Ising Hamiltonian [73]. Any analog quantum device is affected by inherent noise and this noise can lead to minimizing the incorrect Hamiltonian as shown in Sec. 6.1

5. A SEARCH FOR NEW HARDWARE GRAPHS

The advent of analog quantum annealing machines and in particular the D-Wave quantum annealer has sparked a new interest in the study of non-planar Ising spin glasses. While there have been multiple attempts to discern if the D-Wave quantum annealers display an advantage over conventional technologies, to date there are only a few "success stories" where analog quantum optimizers show an advantage over current conventional silicon based computers [15, 21]. Recent results suggest that problems with a more complex energy landscape are needed to discern if quantum annealers can outperform current digital computers [10, 15]. In particular, the search for salient features in the energy landscape, the careful construction of problems with particular features, as well as the attempt to induce a finite-temperature spin-glass transition for lattices restricted to the quasi-two-dimension topologies of the quantum chips have gained considerable attention [93]. In this chapter, we explore two models, the first to reduce the error in finite-size-scaling estimation of critical exponents, and the second to induce a finite-temperature phase transition.

The current generation of quantum annealers is currently limited in the number of qubits with the most recent device, the D-Wave 2000Q, having a maximum of $N = 2048$. However, the estimation of critical exponents requires large systems due to finite-size-scaling effects. One possible method to reduce these scaling corrections is to perform an additional average over graph disorder in addition to the thermal and disorder averaging done in typical estimations of an observable. With this goal, I introduce the bond-diluted next-nearest-neighbor Ising spin-glass. With random dilutions of the graph, a graph disorder average can be performed in an attempt to reduce finite-size-scaling effects.

The quest for a finite-temperature spin-glass transition in quasi-two-dimensional topologies stems from the interest in creating an energy landscape that becomes more complex

and rugged at finite temperatures, such that thermal simulated annealing has a harder time in determining the optimal solution to an Ising-spin-glass-like optimization problem [67]. On the other hand, quantum annealing should, in principle, be able to tunnel through barriers if they are thin enough.

5.1 Scaling Corrections in the Bond-Diluted Next-Nearest-Neighbor Ising Spin Glass

For spin glasses, calculating accurate critical exponents is difficult due to significant corrections to scaling and long equilibration times in Monte Carlo simulations that limit numerical studies to small system sizes. Current quantum annealing devices such as the D-Wave 2 ($N = 512$) and 2X ($N = 1152$) are plagued by the small number of qubits in their respective topologies. In order to study the critical phenomena of such small systems requires a good control of the scaling corrections as emphasized in the work of Hasenbusch *et al.* [94] and Katzgraber *et al.* [3].

One way to reducing finite-size effects in spin glasses is to introduce periodic boundary conditions. Unfortunately, current machines have a fixed connectivity which does not allow open boundaries. The effect of having free boundary conditions is that the edge spins in a topology do not behave like the spins in the bulk. One possible method available to devices with a fixed connectivity is to dilute the bond lattice.

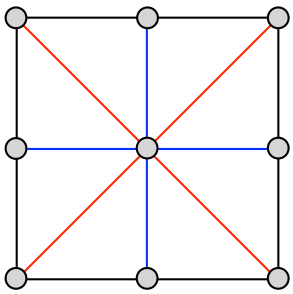

Figure 5.1: Illustration of the two-dimensional next-nearest-neighbor graph. The blue lines are the typical nearest-neighbor interactions. The red lines denote next-nearest-neighbor interactions.

5.1.1 Model

We simulate the three-dimensional next-nearest-neighbor Ising spin glass with coordination number $z = 18$ and Gaussian disorder. Figure 5.1 can be imagined as once slice through the center of a 3×3 cubic lattice. By removing bonds such that each of the spins in the system has an average connectivity of $z = 6$, the same connectivity of the regular three-dimensional cubic lattice, finite size effects should be reduced. Fortunately, a bond-diluted Ising spin glass will share the same universality class with the standard three dimensional cubic lattice [3]. Thus, a comparison of the two models can be performed. Our Hamiltonian is the usual one without a field term;

$$\mathcal{H} = \sum_{\langle i,j \rangle} J_{ij} S_i S_j \qquad \text{with} \qquad S_i \in \{\pm 1\} \tag{5.1}$$

In order to measure corrections to scaling, we measure the Binder cumulant of the overlap and the finite size correlation length respectively,

$$g_q = \frac{1}{2}\left[3 - \frac{\langle q^4 \rangle}{\langle q^2 \rangle^2}\right] \tag{5.2}$$

$$\xi_L = \frac{1}{2\sin(k_{\min}/2)}\left[\frac{\chi_{\mathrm{SG}}(\mathbf{0})}{\chi_{\mathrm{SG}}(k_{\min})}\right] \tag{5.3}$$

where

$$\chi_{\mathrm{SG}} = \frac{1}{N}\sum_{i,j}[\langle S_i S_j \rangle^2] e^{i\boldsymbol{k}\cdot(\boldsymbol{R_i}-\boldsymbol{R_j})} \tag{5.4}$$

and $k_{\min} = (2\pi/L, 0, 0)$, the smallest non-zero wave vector.

5.1.2 Scaling and Corrections

Using the Binder cumulant and correlation length one can approximate the critical exponent ν via

$$g = \tilde{G}(L^{1/\nu}[\beta - \beta_c]) \tag{5.5}$$

$$\frac{\xi_L}{L} = \tilde{X}(L^{1/\nu}[\beta - \beta_c]) \tag{5.6}$$

The spin-glass susceptibility χ_{SG} allows one to estimate the critical exponent η via a knowledge of ν and

$$\chi_{\mathrm{SG}} = L^{2-\eta}\tilde{C}(L^{1/\nu}[\beta - \beta_c]). \tag{5.7}$$

With these scaling equations, all data should fall onto a single curve as shown in Fig. 2.4. Unfortunately, sometimes corrections to scaling are required.

$$g = \tilde{G}(L^{1/\nu}[\beta - \beta_c])[1 + cL^{-\omega} + ...] \tag{5.8}$$

$$\frac{\xi}{L} = \tilde{\xi}(L^{1/\nu}[\beta - \beta_c])[1 + cL^{-\omega} + ...] \tag{5.9}$$

Corrections to scaling are asymptotically dominated by the leading correction to scaling exponent ω and vanish in the thermodynamic limit. Simulation parameters are listed in Table 5.1.

Table 5.1: Parameters of the simulation of the bond-dilute next-nearest-neighbor Ising spin glass. For system size N, we compute $N_{\rm sa}$ instances. $N_{\rm sw} = 2^b$ is the total number of Monte Carlo sweeps for each of the N_T replicas for a single instance, $T_{\rm min}$ [$T_{\rm max}$] is the lowest [highest] temperature simulated, and N_T and $N_{\rm hc}$ are the number of temperatures used in the parallel tempering method and in the isoenergetic cluster algorithm, respectively.

Topology	N	$N_{\rm sa}$	b	$T_{\rm min}$	$T_{\rm max}$	N_T	N_{hc}
3D	64	6000	21	0.700	1.300	13	7
3D	216	6000	21	0.700	1.300	13	7
3D	512	6000	21	0.700	1.300	13	7
3D	1728	3000	21	0.700	1.300	13	7

5.1.3 Results

We compare results from the nearest-neighbor three-dimensional Ising spin glass to the bond-diluted next-nearest-neighbor three-dimensional Ising spin glass. Figures 5.2 and 5.3 show the Binder cumulant and correlation length and their finite-size scaling analysis respectively. Estimates of $\nu_g = 4.2(6)$ and $\nu_\xi = 2.82(9)$ are obtained. For comparison, $\nu_g = 2.67(17)$ and $\nu_\xi = 2.44(9)$ for the three-dimensional nearest-neighbor Ising spin glass. $\nu_\xi \not\approx \nu_g$ implies large corrections to scaling.

After calculating ξ_L/L and g as a function of temperature, we find that it is also useful to plot them as a function of each other. Using Eqs. (5.8) and (5.9), one can write $g = \hat{G}(\xi_L/L)$ where $\hat{G}$ is also a universal scaling function. With this metric, data for all models should collapse onto a single curve. We see in Fig. 5.4 that corrections to scaling are required in order for the data to fall on the same universal curve as the nearest-neighbor data shown in black obtained from Ref. [3].

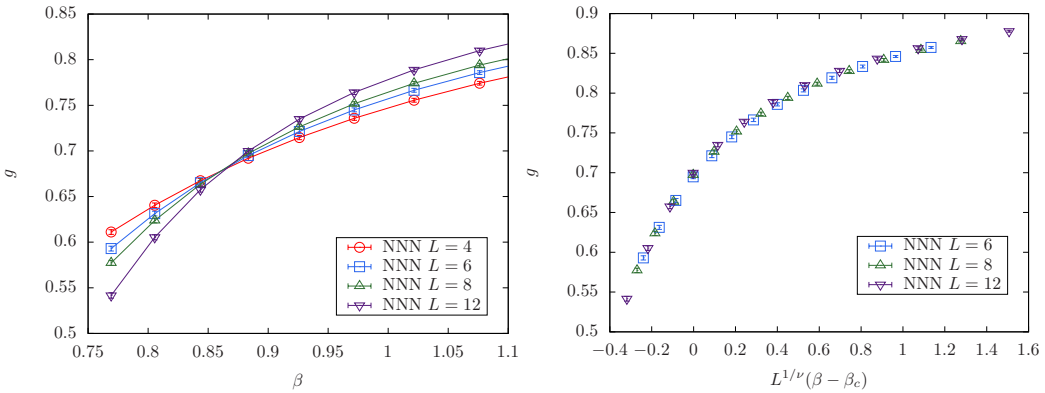

Figure 5.2: Data for the Binder cumulant g_q and its finite-size scaling analysis are shown for the bond-diluted three-dimensional next-nearest-neighbor Ising spin glass. In the finite-size scaling analysis, $\nu = 4.2(6)$ and $T_c = 1.27(5)$. For comparison, $\nu_g = 2.67(17)$ for the three-dimensional nearest-neighbor Ising spin glass.

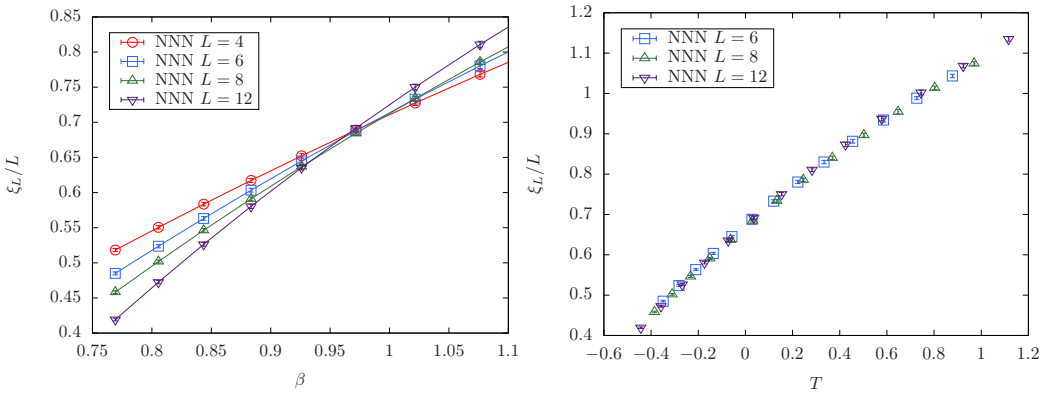

Figure 5.3: Correlation length ξ_L and its finite-size scaling analysis are shown for the three dimensional bond-diluted next-nearest-neighbor Ising spin glass. In the finite-size scaling analysis, $\nu = 2.82(9)$. For comparison, $\nu_\xi = 2.44(9)$ for the three-dimensional nearest-neighbor Ising spin glass. This is a sign of large corrections to scaling.

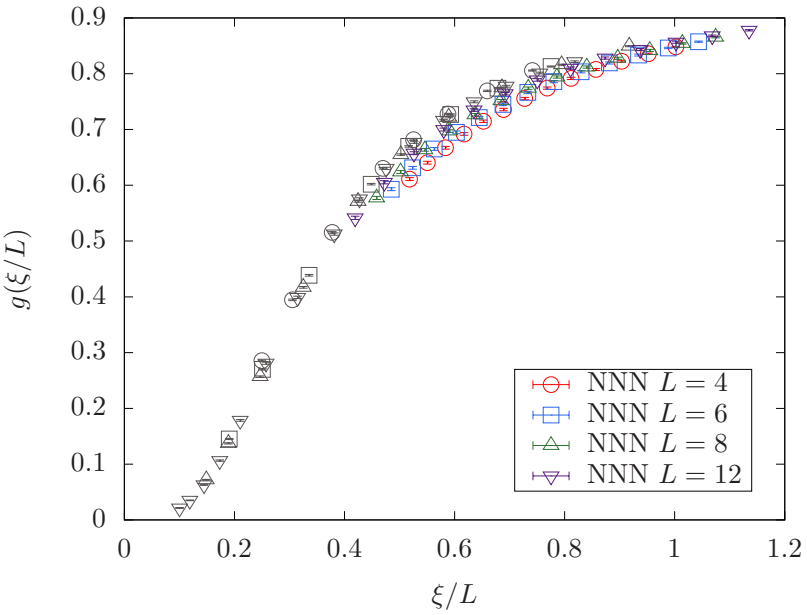

Figure 5.4: Binder cumulant as a function of correlation length three dimensional bond-diluted next-nearest-neighbor Ising spin glass. Data is for the nearest-neighbor three-dimensional Ising spin-glass from Katzgraber *et al.* [3] is shown in black and used as a comparison for the bond-diluted next-nearest-neighbor data. Large corrections to scaling are required to cause the data to fall onto a universal curve.

5.1.4 Summary

By studying the bond-diluted next-nearest-neighbor three-dimensional Ising spin glass we determined this model suffers from large corrections to scaling. This result generates more questions than it answers. The addition of next-nearest-neighbor interactions should not affect the universality [3], however, diluting this next-nearest-neighbor graph produces different critical exponents. It was previously shown by Jörg that the bond-diluted three-dimensional spin glass with Gaussian distributed interactions suffers from large corrections to scaling [95]. There is likely some interplay between the addition of bonds and their random dilution that has not been fully explored.

5.2 Lack of a Finite Temperature Spin-Glass Phase in the Two-Dimensional Randomly Coupled Ferromagnet

The search for problems where quantum annealing might excel over classical optimization techniques has sparked a recent interest in inducing a finite-temperature spin-glass transition in quasi-planar topologies. We study the model of Lemke and Campbell [96], later analyzed in much detail in Refs. [97, 39, 98], that may have the desired finite-temperature spin-glass transition and be of a planar topology that can be easily constructed with current superconducting flux qubits. Unfortunately, our analysis show that for large enough system sizes the model is in a paramagnetic phase at finite temperatures for a parameter range where it is predicted to be a spin glass.

5.2.1 Model

Lemke and Campbell's model is a two-dimensional square-lattice Ising spin glass with uniform ferromagnetic next-nearest-neighbor interactions of strength J and random bi-modal nearest-neighbor interactions of strength $\pm\lambda J$. The Hamiltonian for this model is

$$\mathcal{H} = -\sum_{\langle i,j\rangle} J_{ij} S_i S_j - J \sum_{\langle\langle i,j\rangle\rangle} S_i S_j \qquad \text{with} \qquad S_i \in \{\pm 1\}. \tag{5.10}$$

In Eq. (5.10), $J = 1$ are ferromagnetic interactions between next-nearest-neighbors (denoted by $\langle\langle i,j\rangle\rangle$) and $J_{ij} = \pm\lambda J$ are nearest-neighbor bi-modally distributed spin-glass interactions (denoted by $\langle i,j\rangle$). Depending on the value of λ, Ref. [96] states that a finite-temperature spin-glass transition can be induced in two space dimensions. Extensive numerical simulations by Parisi *et al.* [97] suggested the existence of a crossover in the critical behavior for large enough system sizes. First, from a seemingly ordered state to a spin-glass-like state, followed by a second crossover to a possibly paramagnetic state. This means that true thermodynamic behavior can only be observed if the system sizes

exceed a certain break up length.

However, a conclusive characterization of the critical behavior, as well as the λ-dependence of the break up length ℓ were not discussed in detail until the extensive zero-temperature study by Hartmann and Campbell [98]. By computing ground-state configurations for intermediate system sizes and estimating the stiffness exponent that describes the scaling of energy excitations when a domain is introduced into the system, they argue—based on *zero-temperature* estimates of the spin stiffness—that there should be a finite-temperature spin-glass transition for certain values of λ and linear system sizes L that fulfill $L > \ell$. In particular, they estimate that for $\lambda > \lambda_\infty = 0.27(8)$ no ferromagnetic order should be present. Because the break up length ℓ is large for $\lambda \sim 0.5$ ($L \gtrsim 45$), Ref. [98] suggests studying systems with $\lambda = 0.7$ where $\ell \approx 10$. On the other hand, for $\lambda = 0.90$, the stiffness exponent $\theta = 0.09(5)$ is very close to zero. Therefore, in this work we focus on the cases where (i) we can simulate system sizes $L \gg \ell$ and (ii) the stiffness exponent θ is clearly positive, thus implying a finite-temperature phase, i.e., $\lambda = 0.50$ and 0.75. A summary of the properties of the model for these values of λ, as well as the simulation parameters are listed in Table 5.2.

Table 5.2: Simulation parameters of the two-dimensional randomly coupled ferromagnet and estimates of the stiffness exponent θ and break up length ℓ for different values of λ. For both values of λ we studied different system sizes L using parallel tempering Monte Carlo. The lowest [highest] temperature simulated is $T_{\min} = 0.4$ [$T_{\max} = 2.8$] with $N_T = 50$ temperature steps. Thermalization is tested by a logarithmic binning; once the last three bins agree within error bars we deem the system to be thermalized. For all systems, this was the case after $N_{\rm sw} = 2^{22}$ Monte Carlo sweeps. Furthermore, $N_{\rm sa}$ samples were computed for each parameter combination. Note that the estimate of θ for $\lambda = 0.50$ is taken from Ref. [98], whereas the value for $\lambda = 0.75$ is estimated from the published data (see text for details).

λ	θ	ℓ	L	$N_{\rm sw}$	$T_{\min}$	$T_{\max}$	N_T	$N_{\rm sa}$
0.50	0.59(8)	45	48	2^{22}	0.4	2.8	50	10^4
			64	2^{22}	0.4	2.8	50	10^4
			96	2^{22}	0.4	2.8	50	10^4
			128	2^{22}	0.4	2.8	50	10^4
0.75	0.23(1)	9	24	2^{22}	0.4	2.8	50	10^4
			32	2^{22}	0.4	2.8	50	10^4
			48	2^{22}	0.4	2.8	50	10^4
			64	2^{22}	0.4	2.8	50	10^4

The simulations were performed using parallel tempering Monte Carlo [55] combined with isoenergetic cluster updates [1]. Note that we determine the estimated value of θ for $\lambda = 0.75$ by performing a linear fit to the data of Ref. [98] (quality of fit ~ 0.58 [99]) and estimate $\theta(\lambda) \approx 1.083(3) - 1.12(4)\lambda$, valid in the window $\lambda \in [0.5, 1.1]$. Furthermore, by inspecting Fig. 7 in Ref. [98], we estimate that the break up length for $\lambda = 0.75$ is approximately $\ell \approx 9$.

To detect the existence of a spin-glass transition, we measure the Binder cumulant g

[54] of the spin-glass order parameter q via

$$g_{\rm q} = \frac{1}{2}\left(3 - \frac{[\langle q^4\rangle]_{\rm avg}}{[\langle q^2\rangle]^2_{\rm avg}}\right) \qquad \text{with} \qquad q = \frac{1}{N}\sum_{i=1}^{N} S_i^{\alpha} S_i^{\beta}, \tag{5.11}$$

where α and β represent two copies of the system with the same disorder. The overlap is measured after each Monte Carlo sweep (N spin updates). $\langle ... \rangle$ denotes a thermal average. $[...]$ denote a disorder average over the instances used in the study. The Binder cumulant is dimensionless and scales as $g_{\rm q} = G[L^{1/\nu}(T - T_{\rm c})]$. If $T = T_{\rm c}$, then data for different system sizes cross. If there is no transition, the data does not cross. To rule out a transition at a temperature not simulated a finite-size scaling of the data can be used. To determine what type of phase the simulation is in, we measure the average square of the magnetization

$$m^2 \equiv [\langle m^2\rangle]_{\rm avg} \qquad \text{with} \qquad m = \frac{1}{N}\sum_{i=1}^{N} S_i^{\alpha}. \tag{5.12}$$

5.2.2 Results

We have performed large-scale classical Monte-Carlo simulations for system sizes $L \gg \ell$ and $\lambda = 0.50$ and 0.75. The results for the Binder cumulant which should display a crossing if there is a finite-temperature transition are summarized in Fig. 5.5. the Binder cumulant for the spin-glass order parameter g_q does not show a crossing down to low temperatures for both values of λ studied. In addition, a finite-size scaling of the data for $\lambda = 0.75$ (inset) strongly suggests that $T_c = 0$. Furthermore, the magnetization m^2 as a function of the temperature T decreases with increasing system sizes for both values of λ studied. Based on these results, we conclude that the system is in a *paramagnetic* state for both values of λ in the thermodynamic limit.

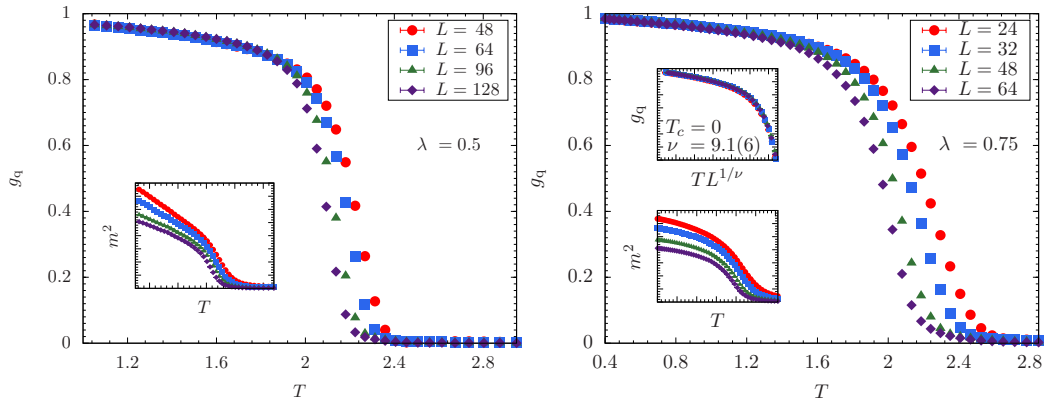

Figure 5.5: Binder cumulant $g_{\rm q}$ for the spin-glass order parameter as a function of the temperature T for the model described in Ref. [96] with $\lambda = 0.50$ (top) and $\lambda = 0.75$ (bottom) and system sizes $L > \ell$. In both cases the data show no crossing at any finite temperature studied, thus suggesting that there is no finite-temperature spin-glass phase. The insets to both panels show that the magnetization m^2 decreases with increasing system size, i.e., the system is likely in a paramagnetic phase. Finally, in the inset we also show a finite-size scaling of the data for $\lambda = 0.75$ that implies $T_c = 0$.

5.2.3 Summary

The results show that the model introduced in Ref. [96] and studied in detail in subsequent publications [97, 39, 98], does not exhibit a finite-temperature spin-glass transition in the thermodynamic limit for values of the parameter λ where it is expected to show such behavior. As suggested in Ref. [97], for large system sizes, the model is in the paramagnetic phase. While Ref. [97] only finds "strong evidence" for the second crossover from spin glass to paramagnet, here we show that the thermodynamic limit is a paramagnetic phase at finite temperature.

Given recent interest in inducing finite-temperature spin-glass transitions in quasi-planar topologies such as the Chimera graph [15], we conjecture that adding any set of interactions that do not grow with the system size to a nearest-neighbor lattice will likely not result in a finite-temperature spin-glass transition [40]. With this in mind, very recent work by Katzgraber *et al.* suggests adding small-world interactions [100] to a chimera

graph, yields a finite-temperature phase transition [101].

6. MITIGATING THE EFFECTS OF NOISE THROUGH CAREFUL CHOICE OF INTERACTIONS*

Despite evidence that the D-Wave quantum annealer indeed runs quantum mechanically, there remain open problems before the device becomes practically useful. One of these problems is the control error, i.e., imperfections in the setting of parameter values of the Ising Hamiltonian in the device [102, 4]. Because it is difficult to set the interactions and local fields of the Hamiltonian with high precision, the device might be attempting to find the ground state of the wrong Hamiltonian, thus compromising the reliability of the final output. This phenomenon arises in any analog device like the D-Wave quantum annealer and it is crucial to devise and implement ingenious methods to mitigate the influence of control errors.

There exist two methods to remedy control errors. First is the reduction of noise on individual qubits and interactions. This has been implemented on the hardware level by D-Wave Inc. for each generation of their quantum annealer. A second method is to perform error-corrected quantum annealing that treats several physical qubits as one logical qubit [103, 104, 105]. However, this method creates overhead by reducing the number of variables in the problem Hamiltonian.

In this section, instead of attempting to correct control errors, I focus directly on working within the noise constraints of analog devices using general methods that can be applied to present and future analog quantum annealers. We present results from classical parallel-tempering Monte Carlo simulations combined with isoenergetic cluster moves from Sec. 4.1 using realistic uncorrelated noise models to study the best-case resilience, i.e., the probability that the ground-state configuration is not affected by random fields and random-bond fluctuations found on the chip. We thus compute classical upper-bound

*Part of this section is reprinted from Ref. [4]. Copyright 2016 by the American Physical Society.

success probabilities.

In addition, we study the discretization of continuous distributions in order to make continuous distributions available to quantum annealing devices with analog noise. The benefits of discretizing continuous distributions extends to special purpose machines such as field programmable gate arrays or FPGAs which have limited memory.

6.1 Benchmarking Resilience to Noise

The recent interest in the D-Wave quantum annealer has sparked a small computing revolution in recent years.[1] In order to discern a quantum advantage to classical algorithms, recent work by Katzgraber *et al.* suggests that tunable hard benchmark problems within the constraints of the D-Wave device is an improvement over spin glasses with uniformly-distributed disorder on the Chimera graph [15]. This work studies the interplay between the generation of hard benchmark instance with the design of problems suitable for the D-Wave device that are robust to noise.

Spin-glasses are incredibly fragile when subjected to small perturbations, also known as chaotic effects, to either couplers (bond chaos), qubits via longitudinal fields (field chaos), or both couplers and qubits by thermal fluctuations (temperature chaos) [106, 107]. We define *resilience*, the probability that the ground-state configuration is not affected by random fields and random-bond fluctuations found on the D-Wave chip for different benchmark instance classes by using realistic uncorrelated noise models for the D-Wave Two quantum annealer. Note that this methodology is generic, i.e., it can be applied to any architecture or noisy black-box optimization device.

6.1.1 Instance Classes

We define a new Hamiltonian on the chimera topology to optimize,

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J_{ij} S_i S_j - \sum_{\langle i \rangle} S_i h_i \qquad \text{with} \qquad S_i \in \{\pm 1\}. \tag{6.1}$$

The disorder distributions are chosen within the hardware constraints of the D-Wave Two architecture. To emulate the effects of thermal noise we perturb the discrete values of the couplers J_{ij} by a random amount ΔJ_{ij} drawn from a Gaussian distribution with mean zero and standard deviation ΔJ. We choose the uncorrelated quenched random fields h in the same manner.

We follow the recipe of Ref. [15] to carefully choose interactions between the spins to determine the hardness and robustness of the instance classes. The ideal benchmark instance is robust to noise, has a unique ground state, and many metastable states. We define a quantity named *yield*, $\mathcal{Y} = N_{\text{unique}}/N_{\text{total}}$. Given N_{total} randomly generated instances, the yield is the ratio of instances with a unique ground state, N_{unique}. We focus our study on the following distributions of interactions which we call instance classes;

- $U_1 \in \{\pm 1\}$
- $U_4 \in \{\pm 1, \pm 2, \pm 3, \pm 4\}$
- $U_{5,6,7} \in \{\pm 5, \pm 6, \pm 7\}$
- $S_{28} \in \{\pm 8, \pm 13, \pm 19, \pm 28\}$.

The $U_{5,6,7}$ and S_{28} instance classes are Sidon sets [108] and reduce the degeneracy of ground states by design, and thus increase yield. Sidon sets are sets of numbers such that all pairwise sums are unique which reduces the chance of having a zero-local field resulting in a "free" spin that increases the degeneracy.

To quantify robustness to noise we define the resilience $\mathcal{R}$ of an individual instance in an instance class to be $\mathcal{R} = N_{\text{same}}/N_{\text{trials}}$ where N_{same} is the number of trials with different

random noise perturbations that do not change the original ground-state configurations. The simulation performs $N_{\text{trials}} = 10$ trials to compute $\mathcal{R}$. Simulation parameters are given in Table 6.1.

Table 6.1: Simulation parameters for benchmarking resilience to noise. For each instance class and system size N, we compute N_{sa} instances. $N_{\text{sw}} = 2^b$ is the total number of Monte Carlo sweeps for each of the $4N_T$ replicas for a single instance, $T_{\min}$ [$T_{\max}$] is the lowest [highest] temperature simulated, and N_T is the number of temperatures used in the parallel tempering method. For the lowest N_{icm} temperatures isoenergetic cluster moves are applied.

Class	N	N_{sa}	b	$T_{\min}$	$T_{\max}$	N_T	N_{icm}
U_1	512	900	19	0.150	3.050	30	13
U_4	512	900	19	0.150	3.000	30	14
$\text{U}_{5,6,7}$	128	900	19	0.150	3.000	30	14
$\text{U}_{5,6,7}$	288	900	19	0.150	3.000	30	14
$\text{U}_{5,6,7}$	512	900	19	0.150	3.000	30	14
$\text{U}_{5,6,7}$	800	900	19	0.150	3.000	30	14
$\text{U}_{5,6,7}$	1152	900	19	0.150	3.000	30	14
S_{28}	512	900	19	0.150	3.000	30	14

6.1.2 Results

We apply the method of fair sampling in Sec. 4.2 to record the configurations that minimize the Hamiltonian and thus, estimate the degeneracy distribution of the ground state. It is important to note that this study only focuses on the resilience of the exact ground state. For the D-Wave Two architecture with 512 qubits, yield is strongly dependent on instance class. U_1 and U_4 have a $\mathcal{Y} = 0$. The yields for $U_{5,6,7}$ and S_{28} are $4.5(4)\%$ and

$20.0(6)\%$ respectively.

Figure 6.1 shows the resilience to random-field and random-coupler noise for the different instance classes. As the strength of the noise increases the resilience decreases due to a high probability of level crossing. Instances with smaller energy gaps, have a lower resilience, again due to an increased chance of level crossing when perturbed by noise.

Recalling that the ideal instance that has a unique ground state, is hard, and robust to noise, some compromise has to be made for the D-Wave Two. Due to the low yield U_1 and U_4 are not useful. S_{28} is too susceptible to noise, while $U_{5,6,7}$ has a non-zero yield and is reasonably robust to noise. The other important conclusion of this work is that the coupler noise has a greater effect on resilience than field noise. In the specific case of $U_{5,6,7}$, Fig. 6.2 shows a dramatic drop in resilience as the system size increases. This means that to scale up the system size of the D-Wave, or any other quantum annealing device, a much more precise control over the device's noise and/or the implementation of error correction schemes [109, 104, 110].

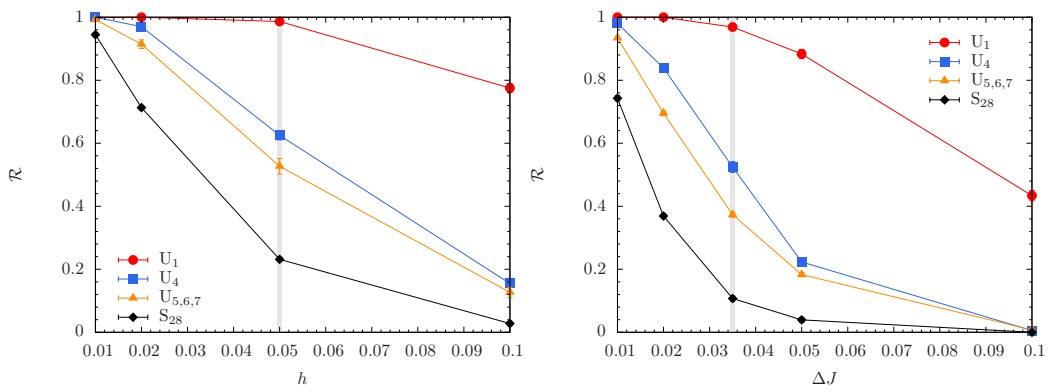

Figure 6.1: Resilience ($\mathcal{R}$) of different instance classes for a $N = 512$ qubit system on the Chimera graph as a function of Gaussian random field strength h (top) and bond fluctuation ΔJ (bottom). Instance classes are less resilient to noise with increasing field strength (bond fluctuation) and decreasing classical energy gap. The shaded line represents the current field (bond fluctuation) noise strength of approximately 5% (3.5%). Note that bond noise has a stronger effect than field noise.

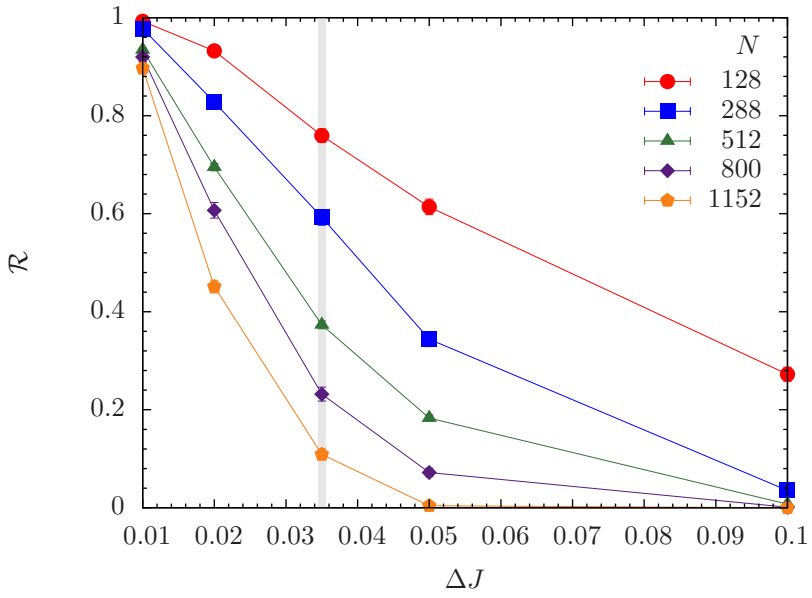

Figure 6.2: Resilience $\mathcal{R}$ of the $U_{5,6,7}$ instance class as a function of the bond fluctuation strength (ΔJ) for different system sizes N on the Chimera topology. The resilience clearly decreases for increasing noise and system size. The shaded vertical line represents the current bond-noise strength in the D-Wave Two system, approximately 3.5%.

Figure 6.3 shows the resilience $\mathcal{R}$ of the $U_{5,6,7}$ instance class as a function of the degeneracy of the first excited state on the Chimera topology with $N = 512$ spins. The higher the degeneracy of the first excited state, the lower the resilience. This can be explained by the increased probability of level crossing. The heat map represents the number of instances that had a given degeneracy N_1 of the first excited state out of the 900 simulated. In this case, the bulk of the instances have between 4 and 8 degenerate first excited states. This results in a reduction of the resilience, compared to instances that contain only one or two first excited states.

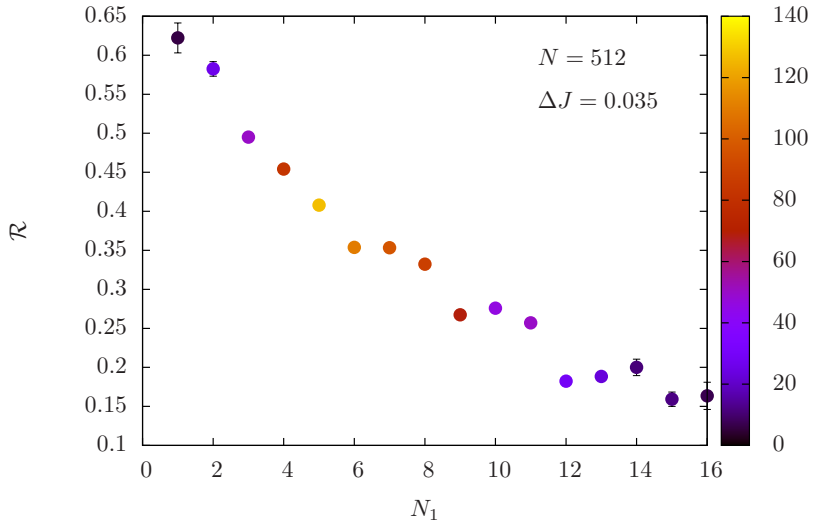

Figure 6.3: Resilience $\mathcal{R}$ as a function of the number of first excited states N_1 for $N = 512$ spins on the Chimera lattice. The data are for the $U_{5,6,7}$ instance class. The color bar shows approximately how often a given number of first excited states occurs for the 900 instances studied. In this case, between four and eight first excited states are most common.

6.1.3 Summary

In order to develop both hard and robust benchmark instances, we tested different instance classes by computing their yield and resilience to noise fluctuations. Ideally, hard instances with a high yield and high resilience are optimal for benchmarking purposes. Both yield and resilience can be tuned by a careful design of the instance classes within the hardware restrictions of the machine followed by a mining of the data. Although the numerical effort to do such mining is non-negligible, this is a key ingredient in designing good benchmarks for quantum annealing devices, as wells as any other computing architectures. It seems that both resilience and yield for the Chimera topology are slightly anti-correlated. A good compromise is thus the $U_{5,6,7}$ instance class where $J_{i,j} \in \pm 5, \pm 6, \pm 7$ that has a good resilience to both field and coupler noise, as wells as a nonzero yield of unique ground states, with a small number of first excited states.

These results for resilience represent a "best-case scenario" for any quantum annealing machine. Any other source of error can only decrease the success probabilities further. However, it could be that carefully crafted correlations between bond and field noise might reduce the error and increase the resilience. Bond noise is the most limiting issue for the D-Wave devices and is highly dependent on the connectivity of the graph. While it is desirable to have a high connectivity to be able to embed interesting problems on any architecture, one has to also keep in mind that noise levels should be far lower than in the current D-Wave machine.

This classical study of both resilience and yield plays an important role in the design of future adjacency matrices for quantum annealing machines, as well as the study of strategies to reduce noise in quantum annealers. The results and methods can easily be generalized to other systems and thus should be of general interest when designing hard instance problems that attempt to circumvent the limitations of current hardware. Calibration of future generations of the D-Wave device should be improved to allow for the encoding of more complex Sidon sets and thus the design of harder benchmark problems. Similarly, although the main goal of this work is to produce problems robust to noise, the methodology can be used to design instances that are particularly sensitive to noise. This could play an important role when designing approaches to better calibrate devices, as done in Ref. [111]. Finally, if either noise is large or the instances produced are too difficult to minimize, a relaxed resilience that includes low-lying excited states can be defined.

6.2 Approximating Continuous Distributions Using Gaussian Quadratures

Special purpose computers, such as the D-Wave 2X quantum annealer or the FPGA-based Janus Computer, are typically restricted by memory constraints, limited precision, or analog noise. This means that the study of problems with interactions drawn from continuous distributions can be difficult on these types of devices. Here we extend the

approach introduced by Leuzzi *et al.* [112] to approximate a continuous Gaussian distribution by using quadratures. Our approach allows us to approximate any continuous distribution using only a few discrete weights. From a classical point of view, this reduces the simulation's memory footprint of continuous problems drastically, as well as the simulation time, because multiple quantities and expensive operations, such as exponentials, can be precomputed and tabulated. For quantum annealing architectures this means that problems that require continuous distributions can be encoded within the restrictions of finite precision and analog noise on these devices.

There are numerous real-world applications which have memory constraints such as the 4-bit precision of the D-Wave Two quantum annealer couplers, as well as special-purpose field programmable gate arrays (FPGAs) utilized in the Janus computer [113]. With the ability to simulate a true Gaussian distribution with a small finite number of discrete values, it is possible to apply more complex algorithms such as the massively parallel population annealing Monte Carlo on memory limited machines such as FPGAs. When calculating the probability to flip a spin in Eq. (3.8), using discretized values of J_{ij} in the Hamiltonian $\mathcal{H}$ allows one to store all the possible values of the update probability in a small memory footprint.

This method was first introduced by Leuzzi *et al.* in Ref. [112], followed by Baity-Jesi *et al.* in Ref. [114] who used the method to discretize the local fields on spins for use in the Janus computer that cannot handle non-integer arithmetic efficiently.

6.2.1 Gauss-Hermite Quadrature

In order to verify thermodynamic quantities we begin with a three-dimensional Ising spin glass with interactions chosen from the continuous normal Gaussian distribution. We then use Gauss-Hermite quadrature [115] to discretize the interactions. Gauss-Hermite

quadrature is useful for approximating integrals of the following kind,

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i) \tag{6.2}$$

where n is the number of nodes used in the discretization and x_i are the roots of the Hermite polynomial with weights w_i given by

$$w_i(n) = \frac{2^{2n-1}(2n)!\sqrt{\pi}}{(2n)^2 H_{2n-1}(x_i)^2} \qquad \text{with} \qquad H_{2n}(x_i) = 0. \tag{6.3}$$

Now, consider a function $f(y)$ where the variable y is normally distributed. The expectation value of f corresponds to the following integral:

$$\langle f(y) \rangle = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2} \right) f(y)dy. \tag{6.4}$$

As this does not correspond to Eq. (6.2), one performs a change of variable $y = \sqrt{2}x$ and Eq. (6.4) becomes

$$\langle f(y) \rangle = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \exp(-x^2) f(\sqrt{2}x)dx \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} w_i f(\sqrt{2}x_i). \tag{6.5}$$

Thus, there are normalization constants $\pi^{-\frac{1}{2}}$ on w_i and $\sqrt{2}$ on x_i. Simulation parameters are given in Table 6.2.

6.2.2 Chebychev-Gauss Quadrature

The method can also be extended to other distributions. One potentially useful distribution is the arcsine distribution, due to its similarities to the bimodal $J_{ij} \in \pm 1$ however with added weight in the center. The recipe for producing nodes and weights is similar however instead of using Hermite polynomials which are useful for approximating inte-

Table 6.2: Parameters of the simulation for approximating continuous distributions with Gaussian quadratures. For each discretization $m = 4, 6, 8$ and 10, and for system size N, we compute $N_{\rm sa}$ instances. $N_{\rm sw} = 2^b$ is the total number of Monte Carlo sweeps for each of the N_T replicas for a single instance, $T_{\rm min}$ [$T_{\rm max}$] is the lowest [highest] temperature simulated, and N_T and $N_{\rm hc}$ are the number of temperatures used in the parallel tempering method and in the isoenergetic cluster algorithm, respectively.

Topology	N	$N_{\rm sa}$	b	$T_{\rm min}$	$T_{\rm max}$	N_T	N_{hc}
3D	216	1000	20	0.2120	1.6325	30	20
3D	512	1000	20	0.2120	1.6325	30	20
3D	1728	1000	20	0.2120	1.6325	30	20

grals of the form e^{x^2}, one uses Chebychev polynomials of the first kind for integrals of the form $(1-x^2)^{-1/2}$. For y chosen from a uniform distribution,

$$\langle f(y)\rangle = \int_{-1}^{+1} \frac{1}{\pi}\frac{f(x)}{\sqrt{1-x^2}}dy \approx \sum_{i=1}^{n} w_i f(x_i) \tag{6.6}$$

where $w_i = 1/n$ with the number of nodes in the discretization n and x_i the zeros of the Chebychev polynomials of the first kind [115]. This distribution is quite easy to simulate due to the equal weights of the nodes. The method of using residuals to determine the best approximation of the probability distribution no longer applies and one can simply increase the number of nodes to improve accuracy.

6.2.3 Results from Classical Simulations

One can calculate the optimal number of nodes n that reproduces the probability distribution and vary the amount of truncation m of that set of nodes. Figure 6.4 shows varying amounts of discretization with the number of nodes. It is apparent that $n = 2$ is most likely insufficient to accurately approximate the normal Gaussian distribution. As a first order approximation, calculating the residual of the weights of each node and the continuous normal probability distribution yields $n = 20$ as the optimum number of nodes to repro-

duce the continuous normal Gaussian distribution. Because the probability of choosing certain nodes falls well below the precision of standard double precision numbers several nodes will simply never be chosen in practice. When considering that the most recent D-Wave 2000Q quantum annealing device has on the order of 10^4 interactions, more nodes can safely be truncated because the probability to choose them will be incredibly small.

Figure 6.5 shows the average energy per spin of the continuous Gaussian distribution as well as its residual. $m = 4$ approximates the continuous Gaussian poorly and is omitted from the plot of the residuals. The approximation improves as the truncation decreases. Measuring the energy is a simple first metric to determine if we can reproduce thermodynamic quantities of the continuous distribution. More sensitive metrics are the higher moments of the spin-glass observable q. In Fig. 6.6, we see that the second moment of the overlap per spin is also well approximated and improves with decreasing truncation.

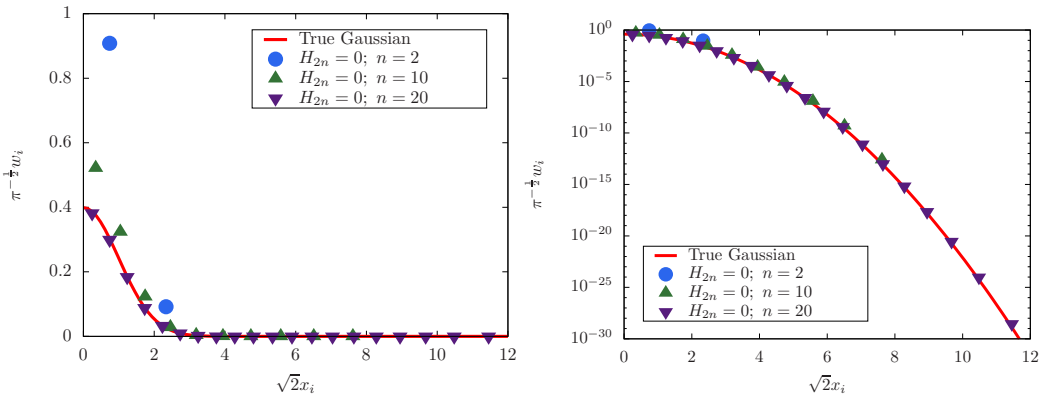

Figure 6.4: Continuous normal Gaussian distribution and different amounts of discretization. The x-axis represents the value of the node in the discretization and the y-axis is the weight or probability with which to choose a particular node. Low orders of discretization greatly overestimate the normal Gaussian probability distribution function. Higher orders of discretization give a very good approximation of the continuous distribution down to 10^{-30}.

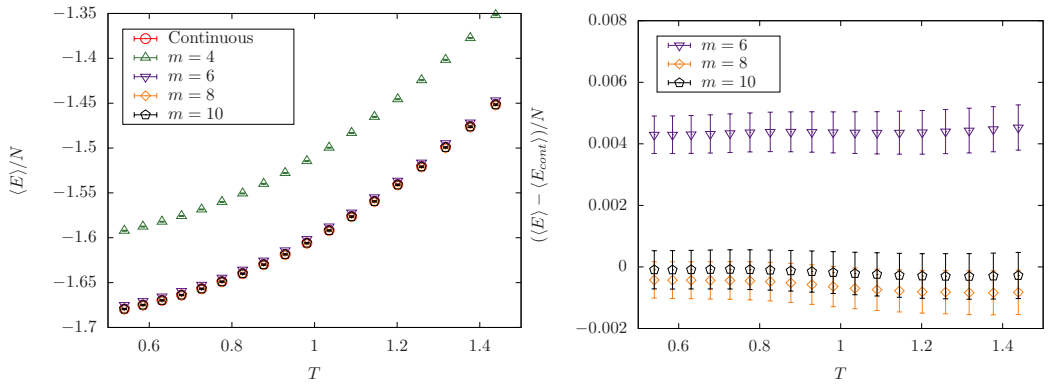

Figure 6.5: Energy and residual energy of the continuous normal Gaussian distribution with $n = 20$ nodes of discretization that has been truncated to m values on a three-dimensional cubic lattice with $N = 1728$. Note that $m = 4$ has been omitted from the plot of residuals due to having error an order of magnitude greater than the other values and is therefore not useful in practical simulations.

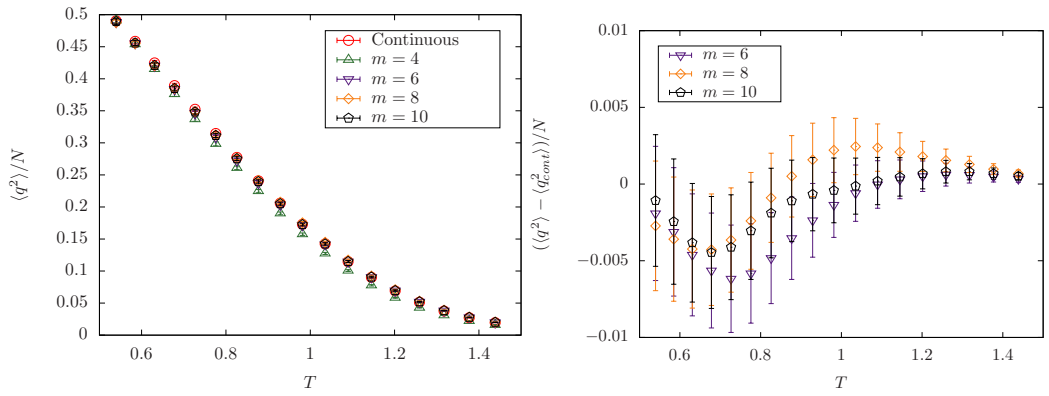

Figure 6.6: Second moment of the overlap its residual of the continuous normal Gaussian distribution with $n = 20$ nodes of discretization that has been truncated to m values on a three-dimensional cubic lattice with $N = 1728$. Note that $m = 4$ has been omitted from the plot of residuals due to having error an order of magnitude greater than the other values and is therefore not useful in practical simulations.

One caveat remains. Due to the discretization, the ground state manifold which only consists of one solution for the continuous Gaussian distribution is now degenerate. Al-

though the discretization may not influence the critical behavior of the system it will influence the low temperature behavior as evidenced by the degenerate ground state solution. Figure 6.7 shows data for the average ground state energy for the continuous normal Gaussian distribution with $n = 20$ nodes of discretization that has been truncated to m values. Although the discretization has increased the degeneracy of the ground-state manifold, the average ground-state energies agree. This is an important feature because discretized distributions can be used to return a similar quality of solution to the cost function.

The same equilibration techniques from Sec. 3.2 are available to discrete Gaussian distributed models. Figure 6.8 shows a measure of equilibration previously applied to continuous Gaussian distributions is also useful for discrete Gaussian models.

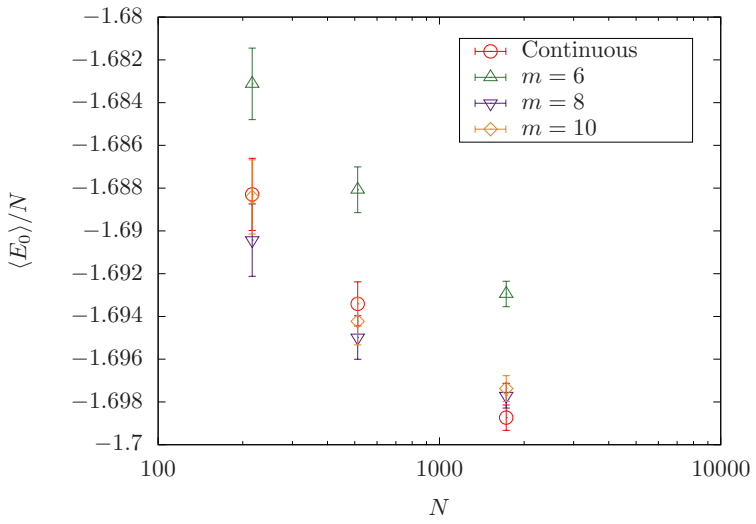

Figure 6.7: Data for the average ground state energy for the continuous normal Gaussian distribution with $n = 20$ nodes of discretization that has been truncated to m values. Although the discretization has increased the degeneracy of the ground-state manifold, the average ground-state energies agree.

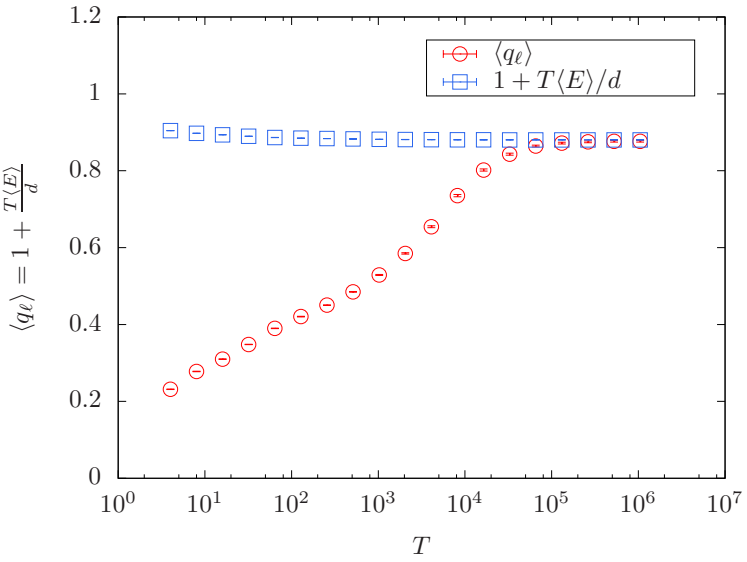

Figure 6.8: Low-temperature, $T \sim 0.2T_c$, equilibration for the three-dimensional Ising spin-glass with N=512 spins and discrete Gaussian interactions with $n = 20$ nodes truncated to $m = 8$. When the two quantities agree, the system is said to be in thermal equilibrium.

Figures 6.9 and 6.10 show measures of $\langle E \rangle / N$ and $\langle q^2 \rangle / N$ and their residuals respectively. The approximations of the continuous arcsine distribution improve as the number of nodes n increases.

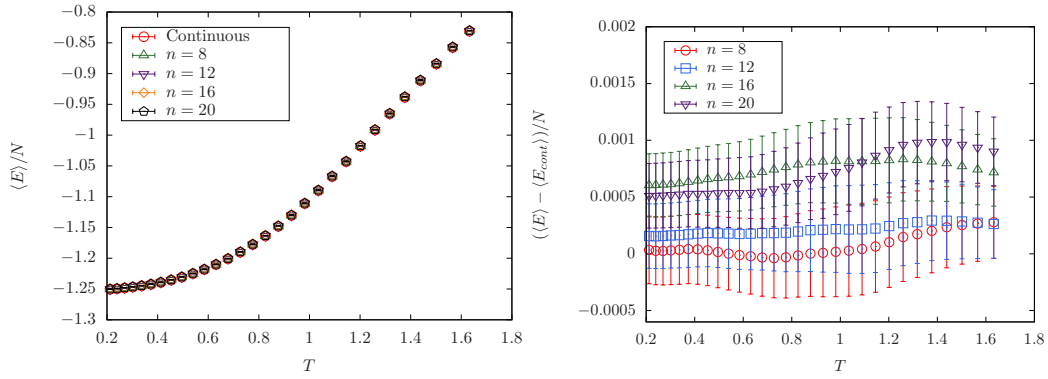

Figure 6.9: Energy and residual energy of the arcsine distribution with n nodes of discretization on a three-dimensional cubic lattices with $N = 1728$.

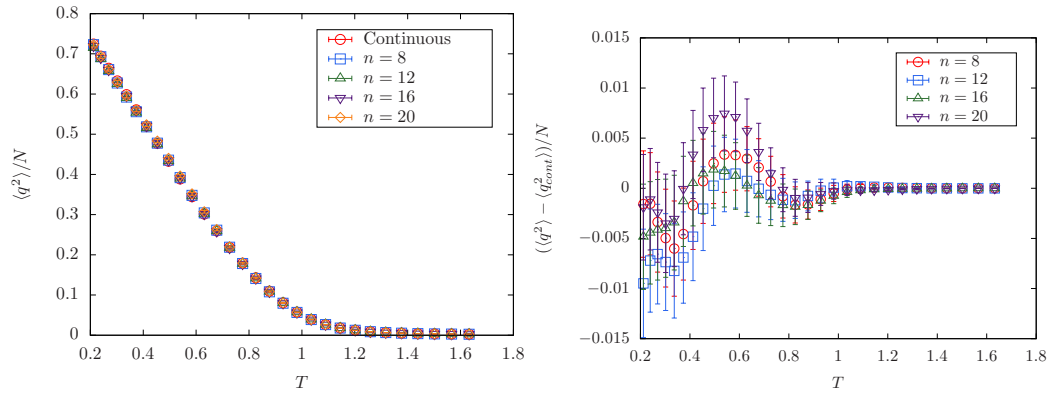

Figure 6.10: Second moment of the overlap its residual of the arcsine distribution with n nodes of discretization on a three-dimensional cubic lattices with $N = 1728$.

6.2.4 Results from Quantum Annealing

Finally, we present results from the D-Wave 2000Q quantum annealing device. In order to quantify improvement we calculate the residual energy ΔE, the difference between the energy found by the D-Wave quantum annealer and the energy found by the isoenergetic cluster algorithm. In Fig. 6.11, the disorder-averaged residual energy from instances on the D-Wave 2000Q is plotted versus system size. For small system sizes the energies agree for both the continuous and Gauss-Hermite quadrature approximation likely because the D-Wave was able to solve both classes of problems. However, for larger system sizes, the Gauss-Hermite approximation produces a smaller residual energy, meaning the interactions chosen from discrete nodes produce a solution closer to the true average ground-state energy.

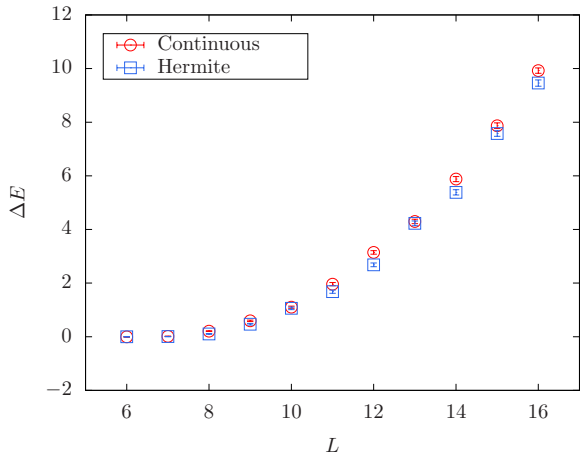

Figure 6.11: Residual energy from simulations on the D-Wave 2000Q ($N = 2048$) quantum annealing device. 100 instances were simulated with each system size and for each instance the residual of the ground-state energy with the true ground-state energy is calculated. For small system sizes the energies agree for both the continuous and Gauss-Hermite quadrature approximation. However, for larger system sizes, the Gauss-Hermite approximation produces a smaller residual energy, meaning the discrete nodes produce a solution closer to the true solution.

6.2.5 Summary

In this section I showed that discretizing continuous distributions of interactions, in this study, the normal Gaussian and arcsine distributions, leads to the same values of thermodynamic observables as the continuous distributions. This discretization allows devices with limited memory such as FPGAs or analog noise such as the D-Wave to simulate previously unaccessible distributions. One side effect is the ground-state manifold is now degenerate, however the average ground-state energy is equivalent within error-bars. This implies solutions are roughly the same quality as those of continuous distributions. The amount of discretization can be tuned and is useful to create instance classes that are more resilient and less susceptible to noise.

The method of Gaussian quadratures can be applied to any distribution of interactions with the careful choice of generating polynomial. This allows any continuous distribution

to be utilized by devices with noise as in the D-Wave quantum annealer, and devices with limited memory which normally would be unable to store transition probabilities otherwise.

7. CONCLUSIONS

7.1 General Summary

I studied several problems to improve the quality and direct the design of current and future quantum annealing devices. In addition, these studies have improved the quality of classical simulations of Ising spin glasses.

In Ch. 2, I introduce the importance of spin glasses in the simulation of complex combinatorial optimization problems such as number partitioning and traveling salesman problems. In addition, the Ising spin glass is the problem the current generation of quantum annealer is designed to solve. With the focus now on spin-glasses, I introduce the necessary physical concepts to understand these simple to state, yet difficult problems. In order to simulate spin glasses, numerical heuristic methods are required due to to the size of the phase space of these problems. In Ch. 3, I introduce the Monte Carlo method, parallel tempering, as well Houdayer cluster moves, as well as a measure of equilibration of spin glasses. An outline of the current relevant algorithms, simulated annealing, quantum annealing, and parallel tempering is given.

After introducing the necessary concepts and methods, in Ch. 4, I present a novel cluster algorithm named *isoenergetic cluster algorithm* to improve the thermalization of spin glasses by restricting Houdayer cluster moves to temperatures below the energy scale of the problem where the interplay of frustration and freezing temperature prevent clusters from spanning the system and becoming inefficient. This algorithm is general in its applicability and can be combined with other optimization methods. Thermalization times are shown to be improved by several orders of magnitude. By including isoenergetic cluster moves in an optimization algorithms, one can find solutions faster than previous state-of-the-art algorithm, in some case by as much as 10^3. This improvement also increases with

the size of the problem. This raises the bar for quantum annealing as an optimizer.

An additional feature of this cluster algorithm is the ability to more fairly sample ground state solutions. I present an algorithm to take advantage of this improved sampling to produce a finite-temperature optimizer that produces unbiased solutions of degenerate problems. A second modification of the algorithm is made to improve generate new solutions from a set of known solutions with little computational effort. One can simply perform cluster moves without the use of Monte Carlo or parallel tempering on known solutions to generate new solutions. This has significant implications in the studies of problems that require many solutions such as satisfiability membership filters as well. This method can also serve as post-processing on the D-Wave quantum annealing device's biased sampling to provide more unique solutions.

Finally we investigated the effect of thermal fluctuations on the ability to infer the ground state in devices with analog noise by measuring the sign of the average spin variable for all spins in a configuration. For system with trivial ferromagnetic solutions, the method works as described by analytic results, however, for non-trivial distributions and solutions, a useful metric for inferring the solution is unavailable. In Sec. 7.2, I describe a possible solution to the problem of decoding spin configurations based on applying thresholds to spin fluctuations.

In Ch. 5, I turn my attention to hardware graphs in order to improve the future design of quantum annealing devices. I begin with the bond-diluted next-nearest-neighbor Ising spin glass in an effort to reduce corrections to scaling by performing graph disorder averaging in addition to thermal and bond-disorder averages. A reduction in corrections to scaling would allow better approximation of critical exponents to characterize novel hardware graphs despite having a smaller number of variables. Unfortunately, this model suffers from large corrections to scaling and the benefits of graph disorder averaging are unclear.

Next I investigate a previously studied model, the two-dimensional randomly coupled

ferromagnet, introduced by Lemke and Campbell [96]. Previous simulations by other parties were uncertain if there exists a finite-temperature spin-glass phase transition for this model. Using numerical values from previous studies, we simulate larger system sizes in the regions of parameter space where a phase transition was thought to exist. However, our results show that the phase transition is a paramagnetic one. Thus, this graph is not useful for future quantum annealing devices that desire a spin-glass phase transition in order to prepare difficult problems for classical simulations and present an opportunity for quantum annealing to excel. However, adding interactions to graphs does show promise as shown in very recent work by Katzgraber and Novotny [101]. In this work, the authors use additional small world bonds to induce a spin-glass phase transition in the quasi-two-dimensional chimera graph where one previously did not exist.

Finally in Ch. 6, I turn my attention to the effects of analog noise in quantum annealing devices. We introduce resilience as a measure of best-case success probability when problem instances are affected by varying amounts of noise. The noise on interactions has a greater effect on the resilience than the noise on the individual spins. With this knowledge it is clear that future manufacturing of quantum annealing devices should prioritize reducing analog noise on qubit interactions.

We also introduce suggestions for how to create classes of problem instances that are robust to noise. These instance classes can be used to generate resilient instances which are less susceptible to noise. It is important for future work to ensure that the desired problem is being solved by the quantum annealer and not a different Hamiltonian due to control errors.

Next we simulate previously unaccessible continuous distributions on systems affected by analog noise. Results from classical simulations show that through the method of Gaussian quadratures, it is possible to discretize continuous distributions, such as the normal Gaussian and arcsine distributions, to reproduce thermodynamic results from the origi-

nal continuous distribution. These discretizations allow special purpose machines such as field programmable gate arrays or graphics processing units which have limited memory to simulate spin-glasses with interactions chosen from continuous distributions.

Summarizing, this body of work focuses on the problem of solving combinatorial optimization problems by simulating spin glasses from three sides: classical algorithm development, suggestions for quantum annealing device design, and improving measurements in realistic physical systems with inherent noise.

7.2 Future Work

Recent work by Karimi *et. al* [116] suggests that if all the solutions found by a heuristic solver are aggregated into a sample, one could ask if there is any additional information in this sample, aside from the solution with the best value. They proceed to develop a method for fixing spins based on information about each spin from the sample and solve the remaining problem which is usually smaller. It would be interesting to apply a similar metric based on thresholds to determine if a spin is in the correct orientation as opposed to a simple majority vote used in the study of finite temperature decoding from Sec. 4.3. Another open question is if it is possible to manipulate the physical finite temperature of quantum annealers in order to allow thermal fluctuations to play a larger role in optimization. It was shown in Chapter 4 that combining sequential algorithms can be useful in improving the thermalization of a simulation.

The overlap of two replicas in a simulation has been used as a measure of "hardness" in developing designer spin-glass instances in order to discover a class of instances that are difficult for classical heuristics due to diverging energy barriers while allowing quantum algorithms to tunnel through these barriers and display remarkable improvements. Katzgraber *et. al* [15] present a recipe for developing these instances through a process of mining random instances. The assumption instances with a histogram of the overlap

with weight near $q = 0$, imply that the two replicas spend their time far from each other in phase space and as a result are more difficult. However, this metric does not provide information as to where the configurations fluctuate. Could it be near or far from the true ground state? I propose a slightly modified metric in which one of the configurations in the overlap is the true ground state which is already known from the mining process. In this case a histogram of the overlap with weight near $q = 1$ implies that the system fluctuates near the ground state whereas weight away from $q = 1$ implies the simulation is far from the ground state in configuration space but not necessarily in energy at low temperatures. This metric could provide a useful determination of hardness for future instance classes.

REFERENCES

[1] Zhu, Z., Ochoa, A. J. & Katzgraber, H. G. Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension. *Phys. Rev. Lett.* **115**, 077201 (2015).

[2] Mandrà, S., Zhu, Z. & Katzgraber, H. G. Exponentially Biased Ground-State Sampling of Quantum Annealing Machines with Transverse-Field Driving Hamiltonians. *Phys. Rev. Lett.* **118**, 070502 (2017).

[3] Katzgraber, H. G., Körner, M. & Young, A. P. Universality in three-dimensional Ising spin glasses: A Monte Carlo study. *Phys. Rev. B* **73**, 224432 (2006).

[4] Zhu, Z., Ochoa, A. J., Hamze, F., Schnabel, S. & Katzgraber, H. G. Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities. *Phys. Rev. A* **93**, 012317 (2016).

[5] Johnson, M. W. *et al.* Quantum annealing with manufactured spins. *Nature* **473**, 194 (2011).

[6] Amin, M. H. S. & Choi, V. First-order quantum phase transition in adiabatic quantum computation. *Phys. Rev. A* **80**, 062326 (2009).

[7] Young, A. P., Knysh, S. & Smelyanskiy, V. N. First-Order Phase Transition in the Quantum Adiabatic Algorithm. *Phys. Rev. Lett.* **104**, 020502 (2010).

[8] Hen, I. & Young, A. P. Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems. *Phys. Rev. E* **84**, 061152 (2011).

[9] Matsuda, Y., Nishimori, H. & Katzgraber, H. G. Ground-state statistics from annealing algorithms: quantum versus classical approaches. *New J. Phys.* **11**, 073021 (2009).

[10] Katzgraber, H. G., Hamze, F. & Andrist, R. S. Glassy Chimeras Could Be Blind to Quantum Speedup: Designing Better Benchmarks for Quantum Annealing Machines. *Phys. Rev. X* **4**, 021008 (2014).

[11] Dickson, N. G. *et al.* Thermally assisted quantum annealing of a 16-qubit problem. *Nat. Commun.* **4**, 1903 (2013).

[12] Boixo, S., Albash, T., Spedalieri, F. M., Chancellor, N. & Lidar, D. A. Experimental signature of programmable quantum annealing. *Nat. Commun.* **4**, 2067 (2013).

[13] Boixo, S. *et al.* Evidence for quantum annealing with more than one hundred qubits. *Nat. Phys.* **10**, 218 (2014).

[14] Rønnow *et al.*, T. F. Defining and detecting quantum speedup. *Science* **345** (2014).

[15] Katzgraber, H. G., Hamze, F., Zhu, Z., Ochoa, A. J. & Munoz-Bauza, H. Seeking Quantum Speedup Through Spin Glasses: The Good, the Bad, and the Ugly. *Phys. Rev. X* **5**, 031026 (2015).

[16] Heim, B., Rønnow, T. F., Isakov, S. V. & Troyer, M. Quantum versus classical annealing of Ising spin glasses. *Science* **348**, 215 (2015).

[17] Hen, I. *et al.* Probing for quantum speedup in spin-glass problems with planted solutions. *Phys. Rev. A* **92**, 042325 (2015).

[18] Rieffel, E. G. *et al.* A case study in programming a quantum annealer for hard operational planning problems. *Quant. Inf. Proc.* **14**, 1 (2015).

[19] Mandrà, S., Zhu, Z., Wang, W., Perdomo-Ortiz, A. & Katzgraber, H. G. Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches. *Phys. Rev. A* **94**, 022337 (2016).

[20] Boixo, S. *et al.* Computational multiqubit tunnelling in programmable quantum annealers. *Nat. Comm.* **7**, 10327 (2016).

[21] Denchev, V. S. *et al.* What is the Computational Value of Finite Range Tunneling? *Phys. Rev. X* **6**, 031015 (2016).

[22] King, J. *et al.* Quantum Annealing amid Local Ruggedness and Global Frustration (2017). (arXiv:quant-phys/1701.04579).

[23] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[24] Nishimori, H. *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).

[25] Finnila, A. B., Gomez, M. A., Sebenik, C., Stenson, C. & Doll, J. D. Quantum annealing: A new method for minimizing multidimensional functions. *Chem. Phys. Lett.* **219**, 343 (1994).

[26] Kadowaki, T. & Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**, 5355 (1998).

[27] Brooke, J., Bitko, D., Rosenbaum, T. F. & Aepli, G. Quantum annealing of a disordered magnet. *Science* **284**, 779 (1999).

[28] Farhi, E., Goldstone, J., Gutmann, S. & Sipser, M. Quantum Computation by Adiabatic Evolution (2000). (arXiv:quant-ph/0001106).

[29] Roland, J. & Cerf, N. J. Quantum search by local adiabatic evolution. *Phys. Rev. A* **65**, 042308 (2002).

[30] Das, A. & Chakrabarti, B. K. *Quantum Annealing and Related Optimization Methods* (Edited by A. Das and B.K. Chakrabarti, Lecture Notes in Physics 679, Berlin: Springer, 2005).

[31] Santoro, G. E. & Tosatti, E. TOPICAL REVIEW: Optimization using quantum mechanics: quantum annealing through adiabatic evolution. *J. Phys. A* **39**, R393 (2006).

[32] Lidar, D. A. Towards Fault Tolerant Adiabatic Quantum Computation. *Phys. Rev. Lett.* **100**, 160506 (2008).

[33] Das, A. & Chakrabarti, B. K. Quantum Annealing and Analog Quantum Computation. *Rev. Mod. Phys.* **80**, 1061 (2008).

[34] Mukherjee, S. & Chakrabarti, B. K. Multivariable optimization: Quantum annealing and computation. *Eur. Phys. J. Special Topics* **224**, 17 (2015).

[35] Lucas, A. Ising formulations of many NP problems. *Front. Physics* **12**, 5 (2014).

[36] Zhu, Z., Ochoa, A. J. & Katzgraber, H. G. Efficient sampling of ground-state configurations for quasi two-dimensional ising spin glasses. *in preparation* (2016).

[37] Mandrà, S. & Katzgraber, H. G. The pitfalls of planar spin-glass benchmarks: Raising the bar for quantum annealers (again). *Quantum Sci. Technol.* **2**, 038501 (2017).

[38] Zhu, Z., Fang, C. & Katzgraber, H. G. borealis - A generalized global update algorithm for Boolean optimization problems (2016). (arXiv:1605.09399).

[39] Lemke, N. & Campbell, I. A. Finite-temperature phase transition in the two-dimensional randomly coupled ferromagnet. *J. Phys. A* **32**, 7851 (1999).

[40] Zhu, Z., Ochoa, A. J. & Katzgraber, H. G. Lack of a thermodynamic finite-temperature spin-glass phase in the two-dimensional randomly-coupled ferromagnet (2016). (arXiv:1604.01796v2).

[41] Nishimori, H. Optimum Decoding Temperature for Error-Correcting Codes. *J. Phys. Soc. Jpn.* **62**, 2973 (1993).

[42] Cook, S. A. The Complexity of Theorem-proving Procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, 151 (ACM, New York, NY, USA, 1971).

[43] Karp, R. M. *Complexity of Computer Computations*, chap. Reducibility among Combinatorial Problems, 85 (New York: Plenum, 1972).

[44] Coppersmith, D. Modifications to the Number Field Sieve. *J. Cryptology* **6**, 169 (1993).

[45] Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comp.* **26**, 1484 (1997).

[46] Ising, E. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31**, 253 (1925).

[47] Onsager, L. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Phys. Rev.* **65**, 117 (1944).

[48] Landau, D. P. & Binder, K. *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, 2000).

[49] Selinger, J. V. *Introduction to the Theory of Soft Matter: From Ideal Gases to Liquid Crystals* (Springer, New York, 2016).

[50] Katzgraber, H. G. Introduction to Monte Carlo Methods (2009). (arXiv:0905.1629).

[51] Edwards, S. F. & Anderson, P. W. Theory of spin glasses. *J. Phys. F: Met. Phys.* **5**, 965 (1975).

[52] Yeomans, J. M. *Statistical Mechanics of Phase Transitions* (Oxford University Press, Oxford, 1992).

[53] Privman, V. (ed.) *Finite Size Scaling and Numerical Simulation of Statistical Systems* (World Scientific, Singapore, 1990).

[54] Binder, K. Critical properties from Monte Carlo coarse graining and renormalization. *Phys. Rev. Lett.* **47**, 693 (1981).

[55] Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **65**, 1604 (1996).

[56] Katzgraber, H. G., Trebst, S., Huse, D. A. & Troyer, M. Feedback-optimized parallel tempering Monte Carlo. *J. Stat. Mech.* P03018 (2006).

[57] Earl, D. J. & Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).

[58] Wolff, U. Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.* **62**, 361 (1989).

[59] Swendsen, R. H. & Wang, J.-S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **57**, 2607 (1986).

[60] Houdayer, J. A cluster Monte Carlo algorithm for 2-dimensional spin glasses. *Eur. Phys. J. B.* **22**, 479 (2001).

[61] Feng, X., Deng, Y. & Blöte, H. W. J. Percolation transitions in two dimensions. *Phys. Rev. E* **78**, 031136 (2008).

[62] Wang, J., Zhou, Z., Zheng, W., Garoni, T. M. & Deng, Y. Bond and site percolation in three dimensions. *Phys. Rev. E* **87**, 052107 (2013).

[63] Katzgraber, H. G., Palassini, M. & Young, A. P. Monte Carlo simulations of spin glasses at low temperatures. *Phys. Rev. B* **63**, 184422 (2001).

[64] De Simone, C. *et al.* Exact ground states in spin glasses: New experimental results with a branch-and-cut algorithm. *J. Stat. Phys.* **80**, 487 (1995).

[65] Pal, K. F. The ground state energy of the Edwards-Anderson Ising spin glass with a hybrid genetic algorithm. *Physica A* **223**, 283 (1996).

[66] Hartmann, A. K. & Rieger, H. *Optimization Algorithms in Physics* (Wiley-VCH, Berlin, 2001).

[67] Kirkpatrick, S., Gelatt, Jr., C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671 (1983).

[68] Isakov, S. V., Kim, Y. B. & Paramekanti, A. Spin-Liquid Phase in a Spin-$1/2$ Quantum Magnet on the Kagome Lattice. *Phys. Rev. Lett.* **97**, 207204 (2006).

[69] Geman, S. & Geman, D. *IEEE Trans. Pattern. Analy. Mach. Intell.* **PAMI-6**, 721 (1984).

[70] Morita, S. & Nishimori, H. Mathematical Foundation of Quantum Annealing. *J. Math. Phys.* **49**, 125210 (2008).

[71] Santoro, G., Martoňák, E., R. Tosatti & Car, R. Theory of quantum annealing of an Ising spin glass. *Science* **295**, 2427 (2002).

[72] Nishimura, K., Nishimori, H., Ochoa, A. J. & Katzgraber, H. G. Retrieving the ground state of spin glasses using thermal noise: Performance of quantum annealing at finite temperatures. *Phys. Rev. E* **94**, 032105 (2016).

[73] Nishimura, K. & Nishimori, H. Quantum annealing with a non-vanishing final value of the transverse field (2017). (arXiv:1708.00236).

[74] Moreno, J. J., Katzgraber, H. G. & Hartmann, A. K. Finding low-temperature states with parallel tempering, simulated annealing and simple Monte Carlo. *Int. J. Mod. Phys. C* **14**, 285 (2003).

[75] Romá, F., Risau-Gusman, S., Ramirez-Pastor, A. J., Nieto, F. & Vogel, E. E. The ground state energy of the Edwards-Anderson spin glass model with a parallel tempering Monte Carlo algorithm. *Physica A* **388**, 2821 (2009).

[76] Wang, W., Machta, J. & Katzgraber, H. G. Comparing Monte Carlo methods for finding ground states of Ising spin glasses: Population annealing, simulated annealing, and parallel tempering. *Phys. Rev. E* **92**, 013303 (2015).

[77] Geyer, C. Monte Carlo Maximum Likelihood for Dependent Data. In Keramidas, E. M. (ed.) *23rd Symposium on the Interface*, 156 (Interface Foundation, Fairfax Station, VA, 1991).

[78] Binder, K. & Young, A. P. Spin Glasses: Experimental Facts, Theoretical Concepts and Open Questions. *Rev. Mod. Phys.* **58**, 801 (1986).

[79] Yucesoy, B., Machta, J. & Katzgraber, H. G. Correlations between the dynamics of parallel tempering and the free-energy landscape in spin glasses. *Phys. Rev. E* **87**, 012104 (2013).

[80] Melchert, O., Katzgraber, H. G. & Novotny, M. A. Site- and bond-percolation thresholds in $K_{n,n}$-based lattices: Vulnerability of quantum annealers to random qubit and coupler failures on chimera topologies. *Phys. Rev. E* **93**, 042128 (2016).

[81] Machta, J. Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes. *Phys. Rev. E* **82**, 026704 (2010).

[82] Gomes, C. P., Sabharwal, A. & Selman, B. Model counting. In Biere, A., Heule, M., van Maaren, H. & Walsch, T. (eds.) *Handbook of Satisfiability* (IOS Press, 2008).

[83] Gopalan, P. *et al.* An FPTAS for # Knapsack and Related Counting Problems. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, 817 (IEEE, Palm Springs CA, 2011).

[84] Weaver, S. A., Ray, K. J., Marek, V. W., Mayer, A. J. & Walker, A. K. Satisfiability-based set membership filters. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)* **8**, 129 (2014).

[85] Wang, W., Machta, J. & Katzgraber, H. G. Evidence against a mean-field description of short-range spin glasses revealed through thermal boundary conditions. *Phys. Rev. B* **90**, 184412 (2014).

[86] Mandrà, S., Zhu, Z. & Katzgraber, H. G. (arXiv:1606.07146.

[87] Wei, W. & Selman, B. A new approach to model counting. In *Theory and Applications of Satisfiability Testing*, 324 (Springer, 2005).

[88] Douglass, A., King, A. D. & Raymond, J. Constructing SAT Filters with a Quantum Annealer. In *Theory and Applications of Satisfiability Testing – SAT 2015*, 104–120 (Springer, Austin TX, 2015).

[89] Ruján, P. Finite Temperature Error-Correcting Codes. *Phys. Rev. Lett.* **70**, 2968 (1993).

[90] Chancellor, N., Szoke, S., Vinci, W., Aeppli, G. & Warburton, P. A. Maximum-Entropy Inference with a Programmable Annealer. *Nat. Sci. Rep.* **6**, 22318 (2016).

[91] Nishimori, H. Internal Energy, Specific Heat and Correlation Function of the Bond-Random Ising Model. *Prog. Theor. Phys.* **66**, 1169 (1981).

[92] Thomas, C. K. & Katzgraber, H. G. Simplest model to study reentrance in physical systems. *Phys. Rev. E* **84**, 040101(R) (2011).

[93] Bunyk, P. *et al.* Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor. *IEEE Trans. Appl. Supercond.* **24**, 1 (2014).

[94] Hasenbusch, M., Pelissetto, A. & Vicari, E. The critical behavior of 3D Ising glass models: universality and scaling corrections. *J. Stat. Mech.* L02001 (2008).

[95] Jörg, T. & Krzakala, F. Comment on "Ultrametricity in the Edwards-Anderson Model" (2007). (arXiv:cond-mat/0709.0894).

[96] Lemke, N. & Campbell, I. A. Two-Dimensional Ising Spin Glasses with Nonzero Ordering Temperatures. *Phys. Rev. Lett.* **76**, 4616 (1996).

[97] Parisi, G., Ruiz-Lorenzo, J. J. & Stariolo, D. A. Crossovers in the two-dimensional Ising spin glass with ferromagnetic next-nearest-neighbour interactions. *J. Phys. A* **31**, 4657 (1998).

[98] Hartmann, A. K. & Campbell, I. A. Ordered phase in the two-dimensional randomly coupled ferromagnet. *Phys. Rev. B* **63**, 094423 (2001).

[99] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1995).

[100] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (2000).

[101] Katzgraber, H. G. & Novotny, M. A. A small-world search for quantum speedup: How small-world interactions can lead to improved quantum annealer designs. *in preparation* (2017).

[102] Perdomo-Ortiz, A., Fluegemann, J., Narasimhan, S., Biswas, R. & Smelyanskiy, V. N. A quantum annealing approach for fault detection and diagnosis of graph-based systems. *Eur. Phys. J., Special Topics* **224**, 131 (2015).

[103] Pudenz, K. L., Albash, T. & Lidar, D. A. Error-corrected quantum annealing with hundreds of qubits. *Nat. Commun.* **5**, 3243 (2014).

[104] Pudenz, K. L., Albash, T. & Lidar, D. A. Quantum Annealing Correction for Random Ising Problems. *Phys. Rev. A* **91**, 042302 (2015).

[105] Young, K. C., Blume-Kohout, R. & Lidar, D. A. Adiabatic quantum optimization with the wrong Hamiltonian. *Phys. Rev. A* **88**, 062314 (2013).

[106] Sasaki, M., Hukushima, K., Yoshino, H. & Takayama, H. Temperature Chaos and Bond Chaos in Edwards-Anderson Ising Spin Glasses: Domain-Wall Free-Energy Measurements. *Phys. Rev. Lett.* **95**, 267203 (2005).

[107] Katzgraber, H. G. & Krzakala, F. Temperature and Disorder Chaos in Three-Dimensional Ising Spin Glasses. *Phys. Rev. Lett.* **98**, 017201 (2007).

[108] Sidon, S. Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen. *Mathematische Annalen* **106**, 536 (1932).

[109] Pudenz, K. L., Albash, T. & Lidar, D. A. Quantum annealing correction for random Ising problems (2014). ArXiv:quant-phys/1408.4382.

[110] Correll, R. R. An Efficient User-Side Nulling Calibration for Quantum Annealing Computers (2015). (arXiv:1503.00700).

[111] Perdomo-Ortiz, A., O'Gorman, B., Fluegemann, J., Biswas, R. & Smelyanskiy, V. N. Determination and correction of persistent biases in quantum annealers. *Nature Sci. Rep.* **6**, 18628 (2016).

[112] Leuzzi, L., Parisi, G., Ricci-Tersenghi, F. & Ruiz-Lorenzo, J. J. Ising Spin-Glass Transition in a Magnetic Field Outside the Limit of Validity of Mean-Field Theory. *Phys. Rev. Lett.* **103**, 267201 (2009).

[113] Belletti, F. *et al.* Simulating spin systems on IANUS, an FPGA-based computer. *Comp. Phys. Comm.* **178**, 208 (2008).

[114] Baity-Jesi, M. *et al.* Dynamical transition in the $D = 3$ Edwards-Anderson spin glass in an external magnetic field. *Phys. Rev. E* **89**, 032140 (2014).

[115] Abramowitz, M. & Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964).

[116] Karimi, H., Rosenberg, G. & Katzgraber, H. G. Effective optimization using sample persistence: A case study on quantum annealers and various Monte Carlo optimization methods (2017). (arXiv:1706.07826).