November 2015

# Population Annealing Monte Carlo Studies of Ising Spin Glasses

wenlong wang
*University of Massachusetts Amherst*

# POPULATION ANNEALING MONTE CARLO STUDIES OF ISING SPIN GLASSES

A Dissertation Presented

by

WENLONG WANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2015

Department of Physics

# POPULATION ANNEALING MONTE CARLO STUDIES OF ISING SPIN GLASSES

A Dissertation Presented

by

WENLONG WANG

Approved as to style and content by:

Jonathan Machta, Chair

Helmut G. Katzgraber, Member

Nikolay Prokofiev, Member

Panayotis G. Kevrekidis, Member

Rory Miskimen, Head
Department of Physics

# ACKNOWLEDGMENTS

First and foremost, I want to thank my professor Jonathan Machta for his kind and patient instructions for many years. It has been an honor to be his student. He has gradually guided me from a student to a more independent researcher. I learned from him to be precise and clear. He helped me kindly while I was in perhaps the hardest time of my life here. I appreciate all his contributions of time, ideas and financial aid to make my PhD experience more interesting and stimulating. He has also impacted me equally on the attitude of life. Being patient and calm in all situations has changed my character a lot and is going to be continuously invaluable to me in the future.

Next, I would like to thank our main collaborator Helmut Katzgraber for helpful discussions and large-scale computational work. Without him, many of the calculations would perhaps never finish. I also want to thank all my committee members for critical comments of my work. I am also in great debt to Nikolay Prokofiev who admitted me into the Physics department and his instructions on quantum many body Physics. I also want to thank Panayotis Kevrekidis for his patient instructions on the study of Bose-Einstein condensates of cold atoms. He has opened a new research area for me on the hot cold atoms.

I want to thank my family and friends who have helped and encouraged me in the past years, especially for my family's great understanding in my studying abroad. They have made my life in America more exciting and meaningful. I also want to thank in particular Dong Yan and Yuping Zhou who have helped a lot in my early years in America and for being my close friends. I also would like to thank Jon's previous student Burcu Yucesoy for helpful discussions and access to her parallel tempering data, which has played an important role during the research. I thank Zhiyuan Yao and Qingyou Meng for simulating discussions and computer assistance.

Finally, I want to thank the United States of America. I enjoy my living here and is experiencing a new culture. It has been a great opportunity to stay and study in the US.

# ABSTRACT

## POPULATION ANNEALING MONTE CARLO STUDIES OF ISING SPIN GLASSES

SEPTEMBER 2015

WENLONG WANG

B.Sc., UNIVERSITY OF SCIENCE AND TECHNOLOGY BEIJING

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jonathan Machta

Spin glasses are spin-lattice models with quenched disorder and frustration. The mean field long-range Sherrington-Kirkpatrick (SK) model was solved by Parisi and displays replica symmetry breaking (RSB), but the more realistic short-range Edwards-Anderson (EA) model is still not solved. Whether the EA spin glass phase has many pairs of pure states as described by the RSB scenario or a single pair of pure states as described by two-state scenarios such as the droplet/scaling picture is not known yet. Rigorous analytical calculations of the EA model are not available at present and efficient numerical simulations of spin glasses are crucial in making progresses in the field. In addition to being a prototypical example of a classical disordered system with many interesting equilibrium as well as nonequilibrium properties, spin glasses are of great importance across multiple fields from neural networks, various combinatorial optimization problems to benchmark tests of quantum annealing machines. Therefore, it is important to gain a better understanding of the spin glass models.

In an effort to do so, our work has two main parts, one is *to develop an efficient algorithm called population annealing Monte Carlo* and the other is *to explore the physics of spin glasses using thermal boundary conditions*. We present a full characterization of the population annealing algorithm focusing on its equilibration properties and make a systematic comparison of population annealing with two well established simulation methods, parallel tempering Monte Carlo and simulated annealing Monte Carlo. We show numerically that population annealing is similar in performance to parallel tempering, each has its own strengths and weaknesses and both algorithms outperform simulated annealing in combinatorial optimization problems.

In thermal boundary conditions, all eight combinations of periodic vs antiperiodic boundary conditions in the three spatial directions appear in the ensemble with their respective Boltzmann weights, thus minimizing finite-size effects due to domain walls. With thermal boundary conditions and sample stiffness extrapolation, we show that our data is consistent with a two-state picture, not the RSB picture for the EA model. Thermal boundary conditions also provides an elegant way to study the phenomena of temperature chaos and bond chaos, and our results are again in agreement with the droplet/scaling scenario.

# TABLE OF CONTENTS

# LIST OF TABLES

**Table** **Page**

# LIST OF FIGURES

**Figure** **Page**

# CHAPTER 1

# INTRODUCTION

I will start with a brief summary of the main ideas of the thesis. The work has two main parts, one is *the development of an efficient algorithm called population annealing*, and the other is *the exploration of the nature of the short-range spin glass phase using thermal boundary conditions*. I will present the characterizations of the population annealing algorithm and compare the efficiency of population annealing with parallel tempering and simulated annealing in spin glass simulations and then use population annealing to explore the nature of the short-range spin glass phase with thermal boundary conditions. In thermal boundary conditions, all eight combinations of periodic vs antiperiodic boundary conditions in the three spatial directions appear in the ensemble with their respective Boltzmann weights, thus minimizing finite-size effects due to domain walls. Both population annealing and thermal boundary conditions can be useful in other disordered statistical mechanical systems. Population annealing has similar performance as parallel tempering but with very different properities and both outperform simulated annealing. Our main conclusion regarding the nature of the short-range spin glass phase is that our data is consistent with the droplet/scaling scenario, not the replica symmetry breaking scenario.

## 1.1 Spin glasses: a general introduction

The prototypical spin glass material are dilute magnetic alloys that have a small concentration (at most a few percent) of magnetic elements like iron or manganese randomly scattered in a host material like a noble metal such as copper or gold [11, 88, 113]. For example, the alloys $Cu_{1-x}Mn_x$ and $Au_{1-x}Fe_x$ were the most extensively studied spin glasses in the early experimental research.

There are also insulating spin glasses such as $Eu_xSr_{1-x}S$, with $x$ roughly between 0.1 and 0.5. Early experimental facts of spin glasses include a cusp in the low-field *ac* magnetic susceptibility at a frequency dependent freezing temperature $T_f$ [22], a rounded maximum of specific heat at a temperature that is slightly higher than the freezing temperature [128], the spins are quenched in time without apparent long-range order, and very long and broadly distributed relaxation and equilibration time scales. From these properties, it was recognized in the early days that the order is neither ferromagnetic nor anti-ferromagnetic, possibly a disordered looking but nevertheless ordered phase with symmetry breaking. However, whether there is a truly disordered phase or the system is simply out of thermal equilibrium like a physical glass was not known. Modern spin glass theories assert that the essential feature of all spin glasses, whether they are metals or insulators, crystals or amorphous solids, is the competition of roughly equally likely ferromagnetic and anti-ferromagnetic interactions between spins. The mechanisms to achieve this can vary from one kind of spin glass to another. In the metallic alloys, the existence of interactions of both signs arises from the Ruderman-Kittel-Kasuya-Yosida (RKKY) interactions, the magnetic dipole-dipole exchange interactions mediated by conduction electrons.

From the perspective of materials science, spin glasses are not so special as alloys or insulators, so why are they interesting then? There are several reasons why understanding spin glasses can be beneficial. *First of all,* spin glasses provide an ideal platform to study *systems with quenched disorder*. Modern condensed matter physics has provided a unified theory of matter from the simultaneous exploration of the first principles of quantum mechanics and statistical mechanics for ordered systems. There are, for example, statistical descriptions for the high temperature thermally disordered phases of liquids and gases, the phonon elementary excitations of lattice vibrations and the Bloch states of electrons and their interactions for the ordered phases of crystals, the theories for the superfluid state of $^4$He and the superconducting state of electrons. But there is so far no unified theory for systems with quenched disorder, like physical glasses. Spin glass is simpler to study than physical glass with a simpler Hamiltonian, therefore spin glass is an ideal system for

the study of disordered systems. Note that disordered systems is a broad class of systems not limited to spin glasses, other well known disordered systems include the random field Ising model and the disorder Bose-Hubbard model. Therefore, disorder systems arises in both classical and quantum statistical physics.

Secondly, spin glasses are related to a wide array of problems in other fields and simulation techniques developed for spin glasses like population annealing Monte Carlo is likely to be useful for other research areas. Spin glasses have been of great importance across multiple fields including condensed matter physics, evolutionary biology, neuroscience and computer science. A wide variety of optimization problems map onto spin-glass-like Hamiltonians [11, 75, 129, 42, 113, 69]. Finding ground states of spin glasses is also one such optimization problem. Spin glasses is also relevant for studying neural networks and protein conformational dynamics [113]. Most recently spin glasses have played a pivotal role in the development of new computing prototypes based on quantum bits in both a theoretical, as well as device-centered role. For example, the stability of topologically-protected quantum computing proposals [39, 63, 18] against different error sources—recently implemented experimentally [87]—heavily relies on spin-glass physics [26, 53, 17]. Similarly, the native benchmark problem currently used to gain a deeper understanding of state-of-the-art quantum annealing machines is based on a spin-glass Hamiltonian [27, 15, 54, 104].

Furthermore, the simultaneous presence of quenched disorder and frustration in spin glasses leads to very complicated free energy landscapes indicating the possibility of metastability, new forms of symmetry breaking, and the emergence of new phenomena such as temperature chaos and bond chaos. Indeed, the symmetry breaking of spin glasses at least in the mean field regime differs from any symmetry breaking in order systems. The extreme sensitivity of the equilibrium spin glass phase to external perturbations such as temperature or the interaction bonds does not present in ordered systems like a ferromagnetic Ising model. Spin glass also display many bizarre non-equilibration phenomena like memory and rejuvenation effect, related to the complicated free energy landscapes.

In addition, spin glasses have been interesting to a relatively new subject called complexity theory [113]. It is also generally accepted that spin glasses are complex systems. The connection of spin glasses to many fields and the emergence of an ultrametric structure of pure states in the mean field theory of spin glasses are reasons why spin glasses are considered as complex systems. The impact of spin glasses on this field continues when it is found that there is a possibility that the mean field theory cannot explain the realistic spin glasses in any finite dimension and the infinite dimension is a singularity. What is the exact consequence of this on statistical physics and complexity theory is however not completely known. If this is true, spin glasses is then indeed a special system as there is no such ordered system as mean field theory generally predicts the essential features of ordering correct and becomes exact above a finite dimension.

In the next sections, we will discuss the spin glass models and the exact solution of a mean field model and our current understanding of short-range spin glasses. We will focus on the equilibrium properties of spin glasses in our work.

## 1.2 Spin glass models

We start by a list of milestones in the history of spin glass theories.

- 1975, Edwards and Anderson [31] proposed a short-range Ising model with quenched random interactions, the Edwards-Anderson (EA) model
- 1975, Sherrington and Kirkpatrick [110] proposed a long-range model, the Sherrington-Kirkpatrick (SK) model
- 1979, Parisi [93, 94, 95, 96] solved the SK model and proposed the replica symmetry breaking (RSB) picture of spin glasses
- 1984-1988, McMillan [74], Bray and Moore [21], Fisher and Huse [35, 36, 37] proposed the droplet/scaling picture for the EA model

- 1996-1998, 2002, Newman and Stein [83] showed that RSB is unlikely to be correct for short-range spin glasses

- 1992-1998 Newman and Stein [82] proposed the chaotic pairs picture

### 1.2.1 The Edwards-Anderson model

The work of Edwards and Anderson (EA) [31] started the modern theory of spin glasses. They proposed that the essential physics of spin glasses is the competition between the quenched ferromagnetic and anti-ferromagnetic interactions. The EA Hamiltonian is

$$H = -\sum_{\langle ij \rangle} J_{ij} S_i S_j - h \sum_i S_i, \tag{1.1}$$

where $S_i$ are the spin degrees of freedom defined on a $d$-dimensional cubic lattice with $S_i = \pm 1$ for Ising spin glasses. The sum over $\langle ij \rangle$ means sum over the nearest neighbour sites. $J_{ij}$ is the coupling between spin $S_i$ and $S_j$, which is chosen from the standard normal distribution $n(0, 1)$ with mean zero and variance one. $h$ is an external magnetic field. There are also other spin glass models like the $\pm J$ model, where $J_{ij}$ is randomly chosen from $\pm 1$. This model has trivial degeneracy in ground states, which is a property of spin glasses that we will look into. The EA model is also more realistic than the $\pm J$ model, therefore, we chose not to study this model in our work. We will refer to each disorder realization of the couplings as a sample. We set $h = 0$ unless specified otherwise and the Hamiltonian has global spin flip symmetry.

The Hamiltonian is similar in form to the regular ferromagnetic Ising model but also with dramatic differences. First of all, the Hamiltonian itself is random with quenched disorder. This is a prototypical example of a class of disordered systems in condensed matter physics. It is interesting to know what is the consequence of such quenched disorder. The experimental origin of the quenched couplings of the EA model comes from the the quenched locations of the magnetic atoms in real spin glass samples. The atoms do not move in experimental, and indeed much longer

time scales, which justifies that the couplings are fixed while the spins undergo thermal fluctuations. Note that there are two types of disorder here, one is quenched disorder and the other is thermal disorder. It is usually the case in disordered systems, one has to do double averages, first a thermal average, and then a disorder average.

It is also not difficult to see that the quenched random ferromagnetic and anti-ferromagnetic interactions lead to frustrations. A frustrated spin loop is shown in Fig. 1.1. $J > 0$ means the interaction is ferromagnetic and $J < 0$ means the interaction is anti-ferromagnetic. If we work from the bottom left spin clockwise at temperature $T = 0$ and assign "up" to the first spin, then we should also have "up" to the second and third to minimize the energy of the system. But the fourth spin has to decide which bond to satisfy. The four bonds cannot be simultaneously satisfied. It is not difficult to see that when the sign of the product of all the $Js$ around a loop is negative, there is frustration for the loop. Spin glass samples typically have a large number of entangled frustrated

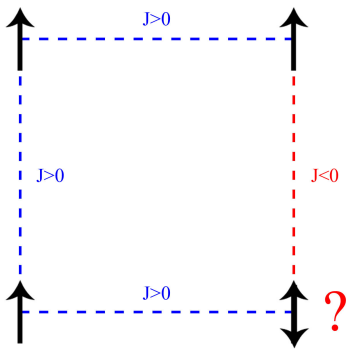


**Figure 1.1.** A frustrated spin loop. $J > 0$ means the interaction is ferromagnetic and $J < 0$ means the interaction is anti-ferromagnetic. Note that no spin configuration can satisfy all the bonds simultaneously.

loops. The simultaneous presence of quenched disorder and frustration causes spin glasses to have

very complicated energy landscapes. The complicated energy landscapes together with the need to average over a large number of samples make spin glass simulations very expensive.

It is worth noting that quenched disorder and frustration are not equivalent in the sense one does not imply the other. One can have quenched disorder without frustration and likewise have frustration without quenched disorder. If we select a random spin configuration and choose the bond connecting two spins as the product of the spin values, then the sample has quenched disorder but no frustration. On the other hand, for the anti-ferromagnetic Ising model on a triangular lattice, there is frustration but no quenched disorder.

Like the mean field theory provides the order parameter for the Ising model, the order parameter of spin glasses can also be found from the mean field solution [88]. Therefore, we will first discuss the mean field spin glass model and its solution in the next section and then come back to the EA spin glass phase afterwards.

### 1.2.2 The Sherrington-Kirkpatrick model

The EA model is difficult to analyze theoretically, so a mean field model was proposed by Sherrington and Kirkpatrick (SK) [110] shortly after the EA model was proposed. The SK Hamiltonian is

$$H = -\frac{1}{\sqrt{N}} \sum_{i<j} J_{ij} S_i S_j - h \sum_i S_i. \tag{1.2}$$

The SK model differs from the EA model in that the interaction is long-range. Each spin interacts with all other spins and there is no geometric structure. The scaling factor $\frac{1}{\sqrt{N}}$ is required to insure that the system has a well-defined thermodynamic limit with a finite energy per spin and also that the first and second terms have the same order of contribution to the energy. Note that the scaling factor differs from that of the long-range Ising model, which is $\frac{1}{N}$. This indicates that even the ground state of the SK model has a comparable number of satisfied and unsatisfied bonds. The interaction energy of each spin scales as $N$ for the Ising model and $\sqrt{N}$ for the SK model. We set $h = 0$ unless specified otherwise and the Hamiltonian has global spin flip symmetry.

The SK model is the most extensively studied and best understood spin glass model. The SK model was solved by Parisi [93, 94, 95, 96] and many features of Parisi's RSB solution of the SK model have now been verified by rigorous mathematical methods [92]. It is known that there is a phase transition at $T_C = 1$. The spin glass phase has a very unusual symmetry breaking called replica symmetry breaking (RSB). A feature of the RSB solution is that instead of a single pair of pure states as is found in the ferromagnetic Ising model, there are a countable infinity pairs of pure states. Only a few pure states have a dominate weight and they have free energy differences of $\mathcal{O}(1)$ and are not related by any symmetry transformation. It is also interesting that these pure states are organized in an ultrametric space. The emergence of such a structure from a structureless SK Hamiltonian is surprising and is an important reason why spin glasses are interesting to the study of complexity.

The spin glass phase transition, like the ferromagnetic phase transition, is second order. The symmetry breaking can be quantified using the concept of order parameters. The order parameter is typically zero in the high temperature symmetric phase, and non-zero in the low temperature symmetry breaking phase. The numerical values of order parameters typically describe the extent of symmetry breaking. Recall that the order parameter of the Ising model is the average magnetization $m$, for a pure state $\alpha$, $m$ is defined as

$$m_\alpha = \frac{1}{N}\sum_{i=1}^{N}\langle S_i\rangle_\alpha, \tag{1.3}$$

where $\langle ...\rangle$ means a thermal average and the subscript $\alpha$ means thermal average within the pure state $\alpha$. In the paramagnetic phase $m = 0$ and in the ferromagnetic phase $m > 0$ or $m < 0$. The two pure states have an infinitely high energy barrier in the thermodynamic limit and transitions of the system from one pure state to the other are dynamically suppressed, hence the name symmetry breaking. The average magnetization $m$ goes continuously from zero to one in the ferromagnetic phase as $T$ changes from the transition temperature $T_C$ to $T = 0$.

The order parameters of spin glasses are quite subtle due to the complicated symmetry breaking, called the spin overlap $q$ distribution. The spin overlap $q$ quantifies the similarity of two pure states. Note that the order parameters are not about particular properties common to all pure states or any single pure state, but rather about the relationships of all pure states. In this sense, the order parameters probe the whole free energy landscape of spin glasses.

The spin overlap $q$ of two pure states $\alpha$ and $\gamma$ is defined as

$$q_{\alpha\gamma} = \frac{1}{N}\sum_{i=1}^{N}\langle S_i\rangle_\alpha\langle S_i\rangle_\gamma. \tag{1.4}$$

A special case of spin overlap is the self-overlap, overlap of a pure state with itself and it is interesting that this quantity is self-averaging and is the same for all pure states. This special quantity is called the Edwards-Anderson order parameter or the EA order parameter $q_{\mathrm{EA}}$. The EA order parameter is therefore defined as

$$q_{\mathrm{EA}} = q_{\alpha\alpha} = q_{\gamma\gamma} = ... \tag{1.5}$$

It is not hard to see that $q_{\mathrm{EA}}$ describes the extent of the spin flip symmetry breaking. The self-averaging of this quantity is in agreement with that the spin glass phase has no apparent long-range order and one pure state looks similar to another. Therefore, $q_{\mathrm{EA}}$ is similar to $m^2$ for the Ising model, and indeed, $q_{\mathrm{EA}}$ depends on temperature and is zero in the paramagnetic phase and increases from zero to one as $T$ decreases from $T_C$ to $T = 0$ in the spin glass phase.

The partition of the phase space to many different pure states allows one to assign a weight to each pure state. Therefore, one can construct a joint distribution using the weights, this is the spin overlap distribution. If there are $M$ pure states, then there are $M^2$ values of $q$, each with a probability as the product of the weights of the two pure states. Of the $M^2$ values of $q$, $M$ of them are equal to $q_{\mathrm{EA}}$ and also $M$ of them are equal to $-q_{\mathrm{EA}}$. What is striking of the SK model is that the $q$ distribution is not self-averaging, the $q$ distribution of all samples have the same $q_{\mathrm{EA}}$ but the

$q$ distribution strictly between $\pm q_{\mathrm{EA}}$ fluctuates from sample to sample, even in the thermodynamic limit. Therefore, the $q$ distribution even though is an important theoretical tool to study spin glasses, does not correspond to any observable in spin glasses.

The idea of pure states only hold in the thermodynamic limit. Therefore, in numerical simulation of finite systems, we do not know which micro-states constitute a pure state. However, this problem can be solved by computing spin overlaps of micro-states $a$ and $b$ as

$$q_{ab} = \frac{1}{N} \sum_{i=1}^{N} S_{ia} S_{ib}, \tag{1.6}$$

where $a, b$ are independently chosen from the Boltzmann distribution. The reason why this works can be understood as follows. Suppose we have two pure states $\alpha$ and $\beta$, the micro-states within each pure state are similar but can be very dissimilar between the two different pure states. The overlap of the two pure states gives a number while the overlap of the two classes of micro-states gives a narrow distribution. It is not hard to show that the mean of the distribution is the same as the overlap of the two pure states. So the overlap distribution computed from micro-states are centred at the correct pure states overlap value but is slightly broadened. Furthermore, the distribution will become narrower and narrower as the number of spins $N$ increases and will indeed converge to the correct pure states overlap value in the thermodynamic limit. This is true for the overlap between any pair of pure states including the self-overlap, and is therefore true for the whole overlap distribution. The overlap distribution computed this way is sufficient to detect the free energy landscape of a sample. In a sense this definition might be even preferred for finite systems as pure states are only well defined in the thermodynamic limit. Three typical $q$ distributions of EA samples of size $L = 8$ at $T = 0.2$ are shown in Fig. 1.2. Note that the $q$ distribution in this case at least for small system sizes varies greatly from one sample to another. The more complicated the $q$ distribution, especially near the origin, the more complicated the energy landscape statistically. Therefore in this case, the blue one is probably the most complicated while the green one and the red one have similar complexity in their energy landscapes.

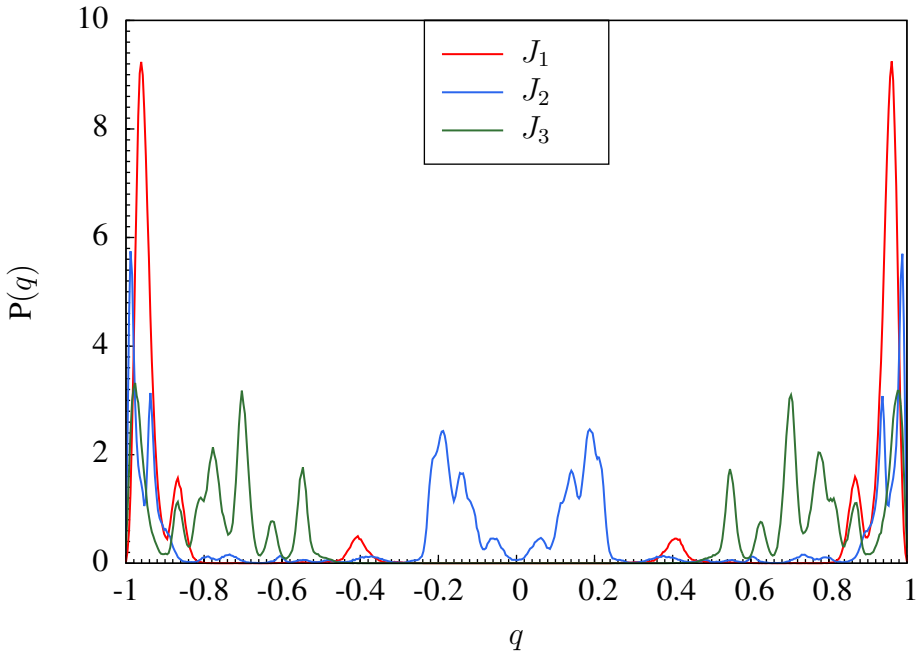


**Figure 1.2.** Three typical $q$ distributions of three-dimensional EA samples of size $L = 8$ at $T = 0.2$. Note that the $q$ distribution fluctuates greatly from sample to sample.

Before we discuss whether RSB can describe the short-range EA model, we summarize the essential results of RSB for the SK model:

1. There are a countable infinity pairs of pure states
2. The $q$ distribution is not self-averaging, but there are always peaks at $\pm q_{\mathrm{EA}}$ for all samples.
3. No overlap strictly between $\pm q_{\mathrm{EA}}$ is special, the disorder averaged $q$ distribution has discrete distributions at $\pm q_{EA}$ with a continuous distribution inbetween.

### 1.2.3 Replica symmetry breaking vs two-state scenarios

Spin glass models are well understood in the mean field regime. However, in finite space dimensions spin glasses are still poorly understood and have been a subject of a long-standing controversy. So, does the EA model behave in the same way as the SK model or does it behave differently from the SK model? Usually, mean field theory provides qualitatively correct symmetry

breaking and becomes exact above some finite dimension. But for spin glasses, the answer may not be so simple.

The first question to answer is whether there is still a phase transition. We know for sure that there is no phase transition in $1d$ and there is a phase transition in infinite dimension. Everything in between is not known exactly, and one has to rely on numerical simulations. There is general agreement that the phase transition in $2d$ is at $T_C = 0$. For $3d$, there is good evidence there is a phase transition at a non-zero temperature. There is more evidence that there is a phase transition in higher dimensions. Binder cumulant [105] and finite size scaling method [23] using the correlation length can be used to determine the transition temperature. The transition temperature for the $3d$ EA model was found to be approximately $T_C \simeq 0.96$ [76] using the correlation length finite size scaling method. If we assume there is indeed a phase transition in $3d$, then the next question is what is the nature of the EA spin glass phase.

The controversy concerning the EA model is between two competing classes of theories as to the nature of the low-temperature phase. One proposal, championed by Parisi and collaborators [93, 95, 96, 100, 75, 129, 98], is that finite-dimensional EA spin glasses behave like the SK model [110]. The mean-field or RSB picture for finite-dimensional EA spin glasses asserts that the qualitative features of the SK model also hold for finite-dimensional models so that, in particular, there are infinitely many pure states in the thermodynamic limit.

In contrast to the RSB picture, the main competing class of theories for the three-dimensional EA model assume that the low-temperature phase consists simply of a single pair of pure states related by the spin-reversal symmetry of the Hamiltonian. The earliest and most widely accepted of these theories is the "droplet picture" developed by McMillan [74], Bray and Moore [19], and Fisher and Huse [35, 36, 37]. The droplet picture asserts that low lying excitations of pure states are compact droplets with energies that scale as a power of the size of the droplet. By contrast, the low lying excitations in the RSB picture are space filling objects.

Several features of the original RSB picture for finite-dimensional EA models have been mathematically ruled out in a series of papers by Newman and Stein [82, 83, 84, 85, 86]. These authors provide two alternative theories for finite-dimensional EA models, both of which have infinitely many pure states. The first is a nonstandard RSB picture, similar to the original RSB picture but with a self-averaging thermodynamic limit. Newman and Stein give heuristic arguments against the nonstandard RSB picture but do not rule it out. On the other hand, the nonstandard RSB picture is promoted as a viable theory for finite-dimensional EA models in Ref. [101]. The second is the "chaotic pairs" picture. Here there are infinitely many pure states but they are organized in such a way that in each finite volume only a single pair of states related by a global spin flip is seen. In our work, we refer to all pictures that display a single pair of pure states in each large finite volume as *two-state* pictures. Therefore, the droplet and chaotic pairs pictures are both two-state pictures within this definition. Note that for the droplet model it is the same pair of states in every volume while for chaotic pairs a different pair of states is manifest in each volume. The difference between RSB and a two-state picture in $q$ distribution is illustrated in Figs. 1.3 and 1.4.

Parallel to these analytical efforts, many computational studies have been aimed at distinguishing between the two classes of theories, see, for example, Refs. [90, 58, 60, 2, 46, 64, 61, 131, 9]. Unfortunately, computational methods have been difficult to apply to spin glasses. The fundamental questions concern the limit of large system sizes, however attempts to extrapolate to large sizes have not been conclusive because the range of sizes accessible to simulations *at low temperatures* is quite small and, for fixed size, the variance between samples for many observables is quite large. Thus, a straightforward extrapolation to large sizes based on mean values of observables can be misleading. Computational studies have yielded a confusing mixture of results, some point to the RSB picture, some to a two-state picture, and some to a mixed scenario, known as the the "trivial nontrivial" (TNT) picture described in Refs. [91, 64, 58]. Recently, there have been efforts to analyze statistics other than simple disorder averages [131, 8, 132, 78, 79, 9] but these methods have not been definitive either and so the controversy continues.

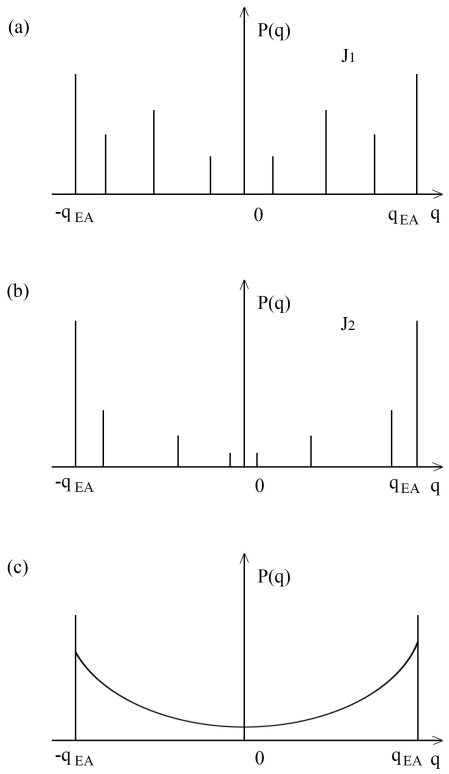


**Figure 1.3.** Distribution of $q$ (a) RSB $J_1$, (b) RSB $J_2$ and (c) disorder average

In the next section, I will summarize some important simulation methods that are used in our work. Many of the methods discussed in the next section can also be useful for other systems with rugged energy landscapes.

## 1.3 Numerical methods

Since the understanding of the EA model in finite dimensions relies heavily on numerical methods, especially Monte Carlo simulations, I therefore give a brief overview in the rest of this chapter of the extensively used simulation methods in our work: population annealing Monte Carlo (PAMC) and parallel tempering Monte Carlo (PTMC) for thermal equilibrium sampling of states, and simulated annealing Monte Carlo (SAMC) for optimization problems. I then discuss

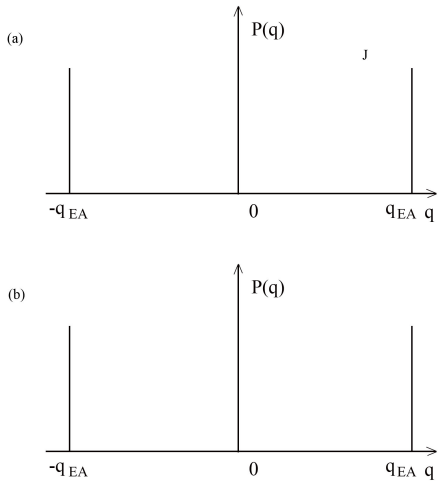


**Figure 1.4.** Distribution of q (a) two-state all $J$, (b) disorder average

the free energy perturbation method [137, 30] for measuring free energy and a useful equilibrium criteria for the EA model. Note that there are also other methods available, for example the Wang-Landau method [121] for measuring density of states, the genetic algorithms [41, 42] and the exact branch-and-cut [111] algorithm for optimization problems.

### 1.3.1 Simulated annealing Monte Carlo

Simulated annealing Monte Carlo (SAMC) [62] is a well known optimization algorithm and is still widely used today in many fields, and it is a generic algorithm. The algorithm starts with a high temperature Markov chain using single temperature Markov chain Monte Carlo (MCMC) algorithms like the Metropolis or the heat bath algorithm and gradually lowers the temperature following an annealing schedule. We used the Metropolis algorithm in our work. Whenever the temperature is lowered, the Metropolis algorithm is applied. This prevents a directly downhill run so that the system has the probability of escaping from local energy minimums. It is beneficial to look at SAMC from the ensemble point of view, since it is usually the case that one has to do many independent runs of SAMC. Furthermore, the comparison between SAMC and PAMC is

apparent in the ensemble picture. As we will see later, the population size of PAMC is the same as the number of independent runs of SAMC with about the same computational work.

The simulated annealing algorithm works as follows:

1. Start with $R_0$ replicas each in thermal equilibrium at a high temperature, either by using random states at the infinite temperature or by using the Metropolis algorithm at a finite but high temperature. Choose an annealing schedule with $N_T$ temperatures. In our implementation, we used a uniform distribution in inverse temperature $\beta$.

2. Lower the temperature by one step following the annealing schedule and then apply $N_S$ Metropolis sweeps independently.

3. Repeat the last step until reaching the lowest temperature of the annealing schedule.

The algorithm is summarized in Fig. 1.5.

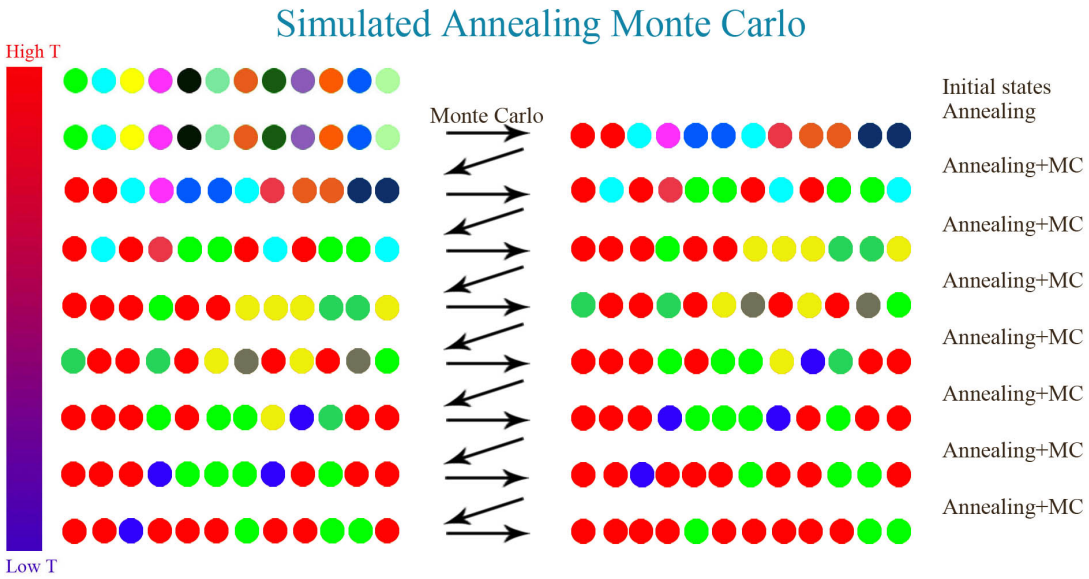


**Figure 1.5.** A schematic diagram of the simulated annealing Monte Carlo algorithm. Each ball of a colour stands for a replica in a micro-state. Replicas are directly copied when temperature is lowered and the metropolis algorithm is applied to all replicas independently.

The ground state can be identified as the lowest energy state encountered in the final ensemble. One can also look for the ground state at all temperatures. It is easy to see that when the energy landscape is sufficiently complicated, simulated annealing will eventually fall out of thermal equilibrium at low temperatures in practice, therefore, simulated annealing is only used for finding ground states of spin glasses and cannot be used for thermal equilibrium sampling of states. We now turn to two algorithms that are efficient in equilibrium sampling of states: population annealing Monte Carlo and parallel tempering Monte Carlo.

### 1.3.2 Population annealing Monte Carlo

We now discuss the population annealing Monte Carlo (PAMC) [48, 70, 71, 125] algorithm. Population annealing is very similar with simulated annealing. Both population annealing and simulated annealing involve taking a population of replicas through an annealing schedule from a high temperature to a low temperature. The difference is that population annealing has an extra resampling step to stay close to thermal equilibrium when the temperature is lowered by doing resampling. When the population size is large and one has a thermally equilibrated ensemble at a temperature $\beta$, then one will have a close to equilibrium ensemble at a nearby lower temperature $\beta'$ after resampling. The resampling step can be viewed as a distribution transformation and greatly improves the efficiency of keeping thermal equilibrium. Note that simulated annealing is thrown out of thermal equilibrium at each annealing step and counts entirely on the Metropolis algorithm to restore equilibrium. Population annealing is simulated annealing with resampling and is a sequential Monte Carlo algorithm.

The population annealing algorithm works as follows:

1. Start with $R_0$ replicas each in thermal equilibrium at a high temperature, either by using random states at the infinite temperature or by using the Metropolis algorithm at a finite but high temperature. Choose an annealing schedule with $N_T$ temperatures. In our implementation, we used a uniform distribution in $\beta$.

2. Lower the temperature of the population with resampling. Suppose the system is cooled from temperature $\beta$ to $\beta^{'}$ with $\beta^{'} > \beta$, the reweighting factor of replica $i$ with energy $E_i$ is proportional to $e^{-(\beta^{'}-\beta)E_i}$ and the expected number of copies of replica $i$ is given by

$$\rho_i(\beta, \beta^{'}) = \frac{e^{-(\beta^{'}-\beta)E_i}}{Q(\beta, \beta^{'})}, \tag{1.7}$$

where Q is the sum of all the re-weighting factors, divided by $R_0$ such that the sum of $\rho_i$ is equal to $R_0$:

$$Q(\beta, \beta^{'}) = \frac{\sum_{i=1}^{R_\beta} e^{-(\beta^{'}-\beta)E_i}}{R_0}, \tag{1.8}$$

where $R_\beta$ is the population size at temperature $\beta$. One then chooses a non-negative integer $n_i$ with expectation value $\rho_i$ and makes $n_i$ copies of replica $i$. There are many ways to do this. The number of replicas can be kept fixed by using the multinomial distribution [71] or the residual resampling method [28]. The number of replicas can also fluctuate by using the Poisson distribution [71] or the nearest integers distribution. We used the nearest integers distribution in our simulations, because it has the smallest variance. The number of copies $n_i$ is either $\lfloor \rho_i \rfloor$ with probability $p = \lceil \rho_i \rceil - \rho_i$ or $\lceil \rho_i \rceil$ with probability $1 - p$.

3. Do MCMC sweeps to all the replicas. Because the new population is now more correlated due to duplications, some MCMC sweeps are needed to decorrelate them. The MCMC sweeps is also necessary to fully explore the phase space. Note that MCMC is only needed for local thermal equilibrium in a potential well, and it is not expected to bring one state from one potential well to another, it is resampling that does this.

4. Repeat the last two steps $N_T - 1$ times to go to the lowest temperature.

The population annealing algorithm is summarized in Fig. 1.6.

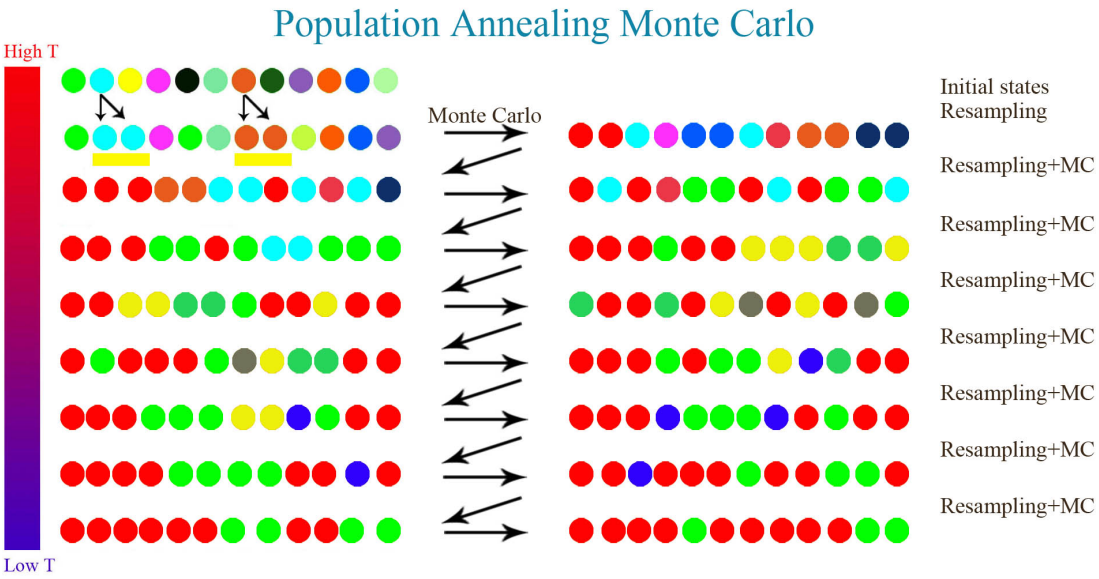


**Figure 1.6.** The population annealing algorithm. Each ball of a colour stands for a micro-state. The metropolis algorithm is applied to all replicas after resampling. Compare with simulated annealing in Fig. 1.5 and note that the only difference of population annealing from simulated annealing is the extra resampling step.

### 1.3.3 Parallel tempering Monte Carlo

Parallel tempering Monte Carlo (PTMC) [114, 40, 49], also known as replica exchange Monte Carlo, is the standard method for simulating spin glasses in the past decades. Parallel tempering simulates an enlarged state space in temperature, and replicas at several different temperatures are simulated in parallel using the Metropolis algorithm. The joint equilibrium of replicas at all temperatures are reached using a swap process that exchanges replicas at two different temperatures. This is usually done by exchanging two replicas at neighbouring temperatures with a swap probability $p_{\text{swap}}$. For two replicas with energy and temperature $E, \beta$ and $E', \beta'$, respectively, the swap probability is

$$p_{\text{swap}} = \min[1, e^{(\beta-\beta')(E-E')}]. \tag{1.9}$$

Note that the swap moves allow a state at a low temperature to diffuse to a high temperature and then go back again. This allows a state at a low temperature to go from one potential well to another indirectly by diffusion. Parallel tempering is a Markov Chain Monte Carlo algorithm.

The parallel tempering algorithm is summarized as follows:

1. Start with $N_T$ replicas at $N_T$ temperatures, ranging from a high temperature to a low temperature.

2. Do Monte Carlo sweeps to all replicas, each sweep includes a regular MCMC sweep to all replicas and $N_T - 1$ swap moves. The swap pairs of nearest neighbours in temperature can be chosen randomly or sequentially.

3. Repeat the last step for $N_S$ sweeps.

The parallel tempering algorithm is summarized in Fig. 1.7.

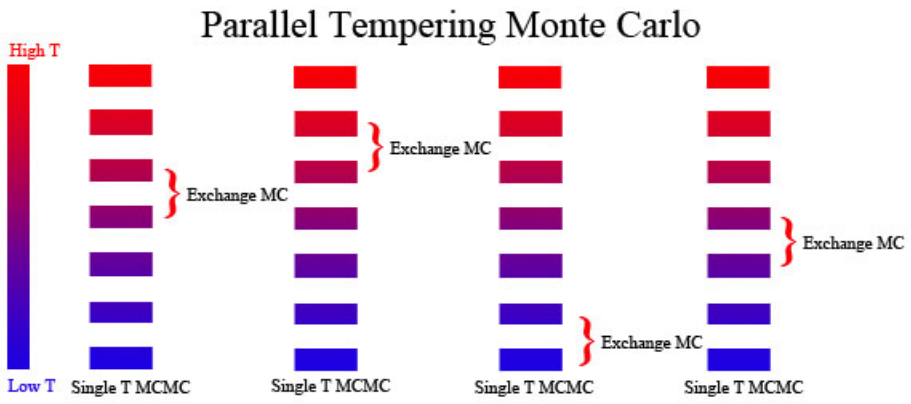


**Figure 1.7.** The parallel tempering algorithm. Each rectangle stands for a micro-state at a certain temperature. The Metropolis algorithm is applied to all replicas at all temperatures in parallel. Besides single temperature Metropolis updates, there are also swap updates which speed up the process of reaching thermal equilibrium for the joint distribution.

### 1.3.4 Thermal boundary conditions

Most Monte Carlo simulations of spin systems are performed with periodic boundary conditions because it is often assumed that periodic boundary conditions yield the mildest finite-size correction. However, for spin glasses with complicated energy landscape, it is likely that periodic boundary conditions may create domain walls that cause finite-size effects. This motivates us to change boundary conditions.

The initial motivation of using thermal boundary conditions starts from the work of Ref. [130], in which it was found for many samples of $L = 8$, when changing boundary conditions, there is almost always a boundary condition that yields a very small weight around $q = 0$ in the overlap distribution. Therefore, we introduced a boundary condition called thermal boundary conditions (TBC). In thermal boundary conditions, the system in each direction has freedom to choose periodic boundary conditions (PBC) or anti-periodic boundary conditions (APBC) according to the Boltzmann weight. For each boundary condition $\zeta$ there is a free energy $F_\zeta$, and the probability distribution for spin states in TBC is the weighted mixture of the eight boundary conditions with weights $e^{-\beta F_\zeta}$. In $3d$, there are a total of 8 boundary conditions that compete with one another. We will investigate the nature of the spin glass phase of the $3d$ EA model using thermal boundary conditions in Chapter. 5.

Thermal boundary conditions also provides an elegant way to study temperature chaos and bond chaos by looking at the scaling properties of quantities at boundary condition crossings and the statistics of boundary condition crossings. We will discuss temperature chaos and bond chaos in Chapter. 6.

### 1.3.5 Free energy perturbation method

Free energy data is important in many statistical mechanical systems. It is interesting that using the equilibrium states of population annealing and parallel tempering, one can accurately measure free energy of spin glasses using the free energy perturbation method [137, 30]. This

method can be used to compute the free energy difference between two nearby temperatures by using the equilibrium states at one of the temperatures.

The ratio of $Z(\beta^{'})$ and $Z(\beta)$, the partition functions at $\beta^{'}$ and $\beta$, respectively, is

$$\frac{Z(\beta^{'})}{Z(\beta)} = \frac{\sum_s e^{-\beta^{'} E_s}}{Z(\beta)} \tag{1.10}$$

$$= \sum_s e^{-(\beta^{'}-\beta)E_s} \left( \frac{e^{-\beta E_s}}{Z(\beta)} \right) \tag{1.11}$$

$$= \langle e^{-(\beta^{'}-\beta)E_s} \rangle_\beta \tag{1.12}$$

$$\approx \frac{1}{R_\beta} \sum_{i=1}^{R_\beta} e^{-(\beta^{'}-\beta)E_i} \tag{1.13}$$

$$= Q_0(\beta, \beta^{'}), \tag{1.14}$$

where the sum over $s$ is sum over all micro-states and the sum over $i$ is sum over all measured states in a Monte Carlo simulation. Take the natural logarithm of both sides,

$$\ln Z(\beta^{'}) - \ln Z(\beta) = \ln Q_0(\beta, \beta^{'}) \tag{1.15}$$

$$-\beta^{'} F(\beta^{'}) = -\beta F(\beta) + \ln Q_0(\beta, \beta^{'}), \tag{1.16}$$

where $F$ is the free energy. At $\beta = 0$, $\ln Z = \ln \Omega$, where $\Omega$ is the total number of micro-states. If we order the temperatures as $\beta_0 < \beta_1 < ... < \beta_k$ and $\beta_0 = 0$, then

$$-\beta_i F(\beta_i) = \sum_{j=0}^{i-1} \ln Q_0(\beta_j, \beta_{j+1}) + \ln \Omega. \tag{1.17}$$

It is straightforward to collect free energy data using population annealing as $Q_0$ is related to the normalization factor $Q$ as $Q_0 = \frac{R_0}{R_\beta} Q$, and it is usually the case that population annealing starts from $\beta = 0$ where the reference free energy is known. Parallel tempering usually does not

use a replica at $\beta = 0$, but, nevertheless, one can also use parallel tempering to measure free energy efficiently. I will discuss how to measure free energy using parallel tempering in Chapter 4.

### 1.3.6 The Katzgraber-Young test

Spin glasses have complex energy landscapes and are difficult to equilibrate, it is important to have sound criteria to ensure thermal equilibrium to draw correct conclusions from the numerical results. Equilibration should be checked for each sample individually for both population annealing and parallel tempering. We will discuss the equilibration measures of both algorithm in greater detail in the next chapter.

One can do an additional check of thermal equilibrium using the Katzgraber-Young criteria [58] for all the samples together. Before showing the relation, we need to define a new quantity called the link overlap $q_l$ between two pure states $\alpha$ and $\gamma$ as

$$q_{l\alpha\gamma} = \frac{1}{N_B} \sum_{\langle ij \rangle} \langle S_i S_j \rangle_\alpha \langle S_i S_j \rangle_\gamma, \tag{1.18}$$

where $N_B$ is the total number of bonds in the system. In $d$ dimensions, $N_B = dN$, where $N$ is the total number of spins of the system.

In practice, the $q_l$ distribution can be computed in a similar way as the $q$ distribution using micro-states $a$ and $b$ defined as

$$q_{l_{ab}} = \frac{1}{N_B} \sum_{\langle ij \rangle} S_{i_a} S_{j_a} S_{i_b} S_{j_b}, \tag{1.19}$$

where $a, b$ are again independently chosen from the correct Boltzmann distribution.

For spin glasses with Gaussian disorder, one can show via integration by parts that at thermal equilibrium,

$$[\langle q_l \rangle] = 1 + \frac{T[\langle u \rangle]}{d}, \tag{1.20}$$

where $u$ is energy per spin and [...] means a disorder average.

It is worth noting that this test works only for boundary conditions with fixed bond configurations like periodic boundary conditions, it doesn't apply to boundary conditions that can fluctuate such as thermal boundary conditions. Furthermore, the test also only works for Gaussian disorder and doesn't work for the $\pm J$ model. Finally, I would like to point out that in our data analysis, we found that the equality can be well satisfied even if some fractions of the samples are out of thermal equilibrium. So the test should be taken as an additional check and should not be considered as a guarantee of thermal equilibration for all samples when the equality is satisfied. Nevertheless, considering there are very few exact relations to test in spin glass simulations, this additional check is still very valuable and is particularly interesting for parallel tempering as it provides a time scale of reaching thermal equilibrium for the algorithm.

## 1.4 Overview

The rest of the thesis is organized as follows: I first give a detailed description of the population annealing algorithm and compare it with parallel tempering in sampling equilibrium states of spin glasses in Chapter 2. I then discuss the performance of population annealing, parallel tempering and simulated annealing for the optimization problem of finding spin glass ground states in Chapter 3 and compare parallel tempering and population annealing in measuring free energy in Chapter 4. Afterwards, I will switch to the study of the properties of the low temperature EA spin glass phase. I will discuss our results of the nature of the $3d$ EA spin glass phase in Chapter 5, temperature chaos and bond chaos in Chapter 6. I summarize and discuss future challenges in Chapter 7.

# CHAPTER 2

# POPULATION ANNEALING MONTE CARLO

Population annealing was introduced by Hukushima and Iba [48]. It is an example of a sequential Monte Carlo method [70] in contrast to parallel tempering, which is a Markov chain Monte Carlo. In this chapter, I will discuss several features of population annealing: weighted averaging, systematic and statistical errors and some details of the implementation of the algorithm. I then compare population annealing and parallel tempering in sampling equilibrium states of spin glasses. This chapter is adapted from Ref. [127].

## 2.1 Weighted averages

Many independent runs of PA for the same system may be combined to reduce both systematic and statistical errors in the measurement of an observable $\mathcal{A}$. Suppose we have carried out $M$ independent runs and obtained estimates $\tilde{\mathcal{A}}_m$, $m = 1, \ldots, M$. Let $\tilde{F}_m(\beta)$ be the free energy estimated in run $m$ at the measurement temperature $1/\beta$. If the different runs have different population sizes, let $R_m$ be the nominal population size in run $m$. Then the best estimator, $\overline{\mathcal{A}}$ for the thermal average of the observable is,

$$\overline{\mathcal{A}} = \frac{\sum_{m=1}^{M} \tilde{\mathcal{A}}_m R_m \exp[-\beta \tilde{F}_m(\beta)]}{\sum_{m=1}^{M} R_m \exp[-\beta \tilde{F}_m(\beta)]}. \tag{2.1}$$

To justify this formula, consider an unnormalized variant of population annealing in which the population is not kept under control but is allowed to grow exponentially. In the resampling step in the unnormalized version of PA, the expected number of copies of replica $j$ is simply the

reweighting factor $\exp\left[-(\beta' - \beta)E_j\right]$. Unnormalized PA is equivalent to standard PA except that it requires exponential computer resources and yields better statistics. Without the normalization factor in the resampling step, each replica evolves independently and combining separate runs of the unnormalized algorithm requires no weighting factor other than the obvious weighting by the population size, $R_m$. Thermal averages in unnormalized PA are obtained using simple averaging. The simple average in unnormalized PA becomes a weighted average in standard PA because the populations in different runs of standard PA have been normalized differently. Specifically, the product of the normalization factors $Q$, from the highest temperature to the measurement temperature is the ratio of the population size in unnormalized PA to the population size in standard PA. But this product is proportional to the exponential of the free energy, justifying the use of $R_m \exp[-\beta \tilde{F}_m(\beta)]$ as the weighting factor in standard PA. Observables such as the spin and link overlap that involve more than one independent copy of the system may also be estimated using weighted averages from multiple independent runs as discussed below in Sec. 2.3.3.

Weighted averaging for the dimensionless free energy is more complicated because the free energy involves measurements at all temperatures however, as shown in Ref. [70], the final results is relatively simple,

$$-\beta\overline{F} = \log\left[\frac{\sum_{m=1}^{M} R_m \exp[-\beta\tilde{F}_m]}{\sum_{m=1}^{M} R_m}\right]. \tag{2.2}$$

This equation is obtained from the fact that $Q(\beta_\ell, \beta_{\ell+1})$ is an observable for which weighted averaging applies, but at inverse temperature $\beta_\ell$. Thus

$$\begin{aligned} &-\beta_k\overline{F}(\beta_k) = \\ &\sum_{j=0}^{k-1} \log\left[\frac{\sum_{m=1}^{M} Q_m(\beta_j, \beta_{j+1}) R_m \exp[-\beta_j\tilde{F}_m(\beta_j)]}{\sum_{m=1}^{M} R_m \exp[-\beta_j\tilde{F}_m(\beta_j)]}\right] \\ &\quad + \log\Omega. \end{aligned} \tag{2.3}$$

This complicated equation for the weighted average of the dimensionless free energy collapses to Eq. 2.2 after using the fact that

$$Q_m(\beta_\ell, \beta_{\ell+1}) \exp[-\beta_\ell \tilde{F}_m(\beta_\ell)] = \exp[-\beta_{\ell+1} \tilde{F}_m(\beta_{\ell+1})], \tag{2.4}$$

and also noting that the weighting factor at $\beta = 0$ is simply $R_m$ and setting $\beta_k = \beta$.

It is important to understand that combining multiple independent runs with weighted averaging reduces both statistical errors *and* systematic errors. By contrast, ordinary averaging reduces only statistical errors. It is obvious that more measurements should reduce statistical errors. Systematic errors are reduced because the weighted average of multiple runs is identical to simulating a larger population size and systematic errors diminish with population size. Indeed, all ensemble averaged quantities are *exact* in the limit of an infinite population size or, equivalently, using weighted averaging in the limit of an infinite number of runs with fixed population size. If the variance in $\beta\tilde{F}(\beta)$ is much less than unity, there is little difference between weighted averaging and simple averaging. However, if the variance of the free energy is large, the weighting factors, which depend exponentially on the free energy, are broadly distributed, and the two averages differ substantially. As we shall see in the next subsection, the variance of the free energy estimator is a fundamental quantity in understanding systematic errors in PA.

There is no method available for combining independent runs of a MCMC algorithm to decrease systematic errors. The most comparable procedure to weighted averaging for MCMC algorithms is 'checkpointing.' In checkpointing, the complete state of the system is saved at the end of the simulation. If results with smaller systematic errors are required, the simulation can be re-started beginning with the final state of the previous simulation so that averaging is initiated after a longer initialization period. Compared to weighted averaging, checkpointing requires substantially more storage since the full configuration of the system must be stored, instead of just the estimators for the observables and the free energy. In addition, checkpointing must be done sequentially while weighted averaging can be carried out using multiple parallel runs. It is a significant advantage of PA that independent runs can be combined to improve systematic errors (equilibration), which is not possible in PT.

## 2.2 Systematic and statistical errors

### 2.2.1 Systematic errors and the variance of the free energy

Systematic errors in PA reflect the fact that for finite population size $R$, the population is not an unbiased sample from the Gibbs distribution. For PA, the algorithm 'equilibrates' to the Gibbs distribution as $R$ increases. In this section we study the convergence in $R$ to the equilibrium Gibbs distribution. Consider the weighted average of $M$ runs each with fixed population size $R$. In what follows a fixed value of $R$ is implicit in the notation. We argued in Sec. 2.1 that the exact Gibbs ensemble average $\langle \mathcal{A} \rangle$ of observable $\mathcal{A}$ is obtained by weighted averaging in the limit of infinitely many runs,

$$\langle \mathcal{A} \rangle = \lim_{M \to \infty} \frac{\sum_{m=1}^{M} \tilde{\mathcal{A}}_m \exp[-\beta \tilde{F}_m(\beta)]}{\sum_{m=1}^{M} \exp[-\beta \tilde{F}_m(\beta)]}. \tag{2.5}$$

Replacing the sum over runs by an integral over classes of runs, we obtain,

$$\langle \mathcal{A} \rangle = \frac{\int \int x \; p_{\mathcal{A}F}(x, y) \exp[-\beta y] dx dy}{\int p_F(y) \exp[-\beta y] dy}, \tag{2.6}$$

where $p_F(\cdot)$ is the probability density for free energy estimator $\tilde{F}$ and $p_{\mathcal{A}F}(\cdot, \cdot)$ is the joint probability density of measuring observable $\tilde{\mathcal{A}}$ and free energy estimator $\tilde{F}$. The average of the estimator $\tilde{\mathcal{A}}$ in a single run of PA, $\langle \tilde{\mathcal{A}} \rangle$ is

$$\langle \tilde{\mathcal{A}} \rangle = \int x \; p_{\mathcal{A}}(x) dx. \tag{2.7}$$

Note that the difference between the integrals for $\langle \mathcal{A} \rangle$ and $\langle \tilde{\mathcal{A}} \rangle$ is simply the weighting factor $\exp[-\beta \tilde{F}]$. The difference, $\Delta \mathcal{A} = \langle \tilde{\mathcal{A}} \rangle - \langle \mathcal{A} \rangle$ is the systematic error in measuring $\mathcal{A}$ in a single run of PA with population size $R$.

Systematic errors for the free energy present a simpler situation. The cumulant generating function $\phi$ of $p_F$ is defined as,

$$\phi(z) = \log \left[ \int dy \exp[zy] \, p_F(y) \right]. \tag{2.8}$$

But $\phi(-\beta)$ is the integral expression for the weighted average of the dimensionless free energy, see Eq. 2.2 with constant $R_m$. Thus, the equilibrium free energy $F$ is related to the distribution of the free energy estimator via,

$$F = -\phi(-\beta)/\beta, \tag{2.9}$$

while the expected value of the free energy estimator from a single run, $\langle \tilde{F} \rangle$, is given by

$$\langle \tilde{F} \rangle = \frac{\partial}{\partial z} \phi(z) \bigg|_{z=0} = \mu_F \tag{2.10}$$

where $\mu_F$ is the mean of $p_F$. The systematic error in the free energy is the difference between these expressions, $\Delta F = \langle \tilde{F} \rangle - F$. Since $\phi(z)$ is the cumulant generating function, we see that

$$\Delta F = \frac{1}{2} \beta \sigma_F^2 + \sum_{n=3}^{\infty} \frac{(-1)^n \beta^{n-1}}{n!} C_n, \tag{2.11}$$

where $C_n$ is the n$^{th}$ cumulant of $p_F$ and $\sigma_F^2 = C_2$ is the variance of $p_F$.

For large population size, $R \gg 1$ a central limit theorem argument suggests that $p_F$ should become a Gaussian since $\tilde{F}$ is the sum of contributions from large number of nearly independent members of the population. Thus for large $R$ we expect the simpler expression,

$$\Delta F = \frac{\beta}{2} \sigma_F^2 \tag{2.12}$$

to become exact.

Similarly, for large $R$ we expect the joint distribution $p_{\mathcal{A}F}$ in Eq. 2.6 to be a bi-variate Gaussian defined by the means and variances of $\tilde{\mathcal{A}}$ and $\tilde{F}$, and their covariance, $\mathrm{cov}(\tilde{\mathcal{A}}, \tilde{F})$. Carrying out the Gaussian integrals for $\langle \mathcal{A} \rangle$ in Eq. 2.6 we obtain for the equilibrium value of the observable,

$$\langle \mathcal{A} \rangle = \mu_A - \beta \, \mathrm{cov}(\tilde{\mathcal{A}}, \tilde{F}). \tag{2.13}$$

Thus the systematic error $\Delta\mathcal{A} = \langle\tilde{\mathcal{A}}\rangle - \langle\mathcal{A}\rangle$ in estimating the observable $\mathcal{A}$ with population size $R$ is given by

$$\Delta\mathcal{A} = \beta\,\mathrm{cov}(\tilde{\mathcal{A}}, \tilde{F}). \tag{2.14}$$

We see that for large $R$ the systematic error in any observable is proportional to the covariance of the observable with the free energy estimator. This expression for the systematic error in $\mathcal{A}$ can be re-written in a form that emphasizes the central role of the variance of the free energy,

$$\Delta\mathcal{A} = \mathrm{var}(\beta\tilde{F}) \left[\frac{\mathrm{cov}(\tilde{\mathcal{A}}, \beta\tilde{F})}{\mathrm{var}(\beta\tilde{F})}\right]. \tag{2.15}$$

It is expected that the quantity in the square bracket will be nearly independent of $R$ so that systematic errors in $\mathcal{A}$ are proportional to the variance of dimensionless free energy, just as is the case for the free energy itself.

A central limit theorem argument suggests that $\mathrm{var}(\beta\tilde{F})$ decreases as $1/R$ so that the product $R\,\mathrm{var}(\beta\tilde{F})$ should approach a constant. Define the *equilibration population size*, $\rho_f$ as

$$\rho_f = \lim_{R\to\infty} R\ \mathrm{var}(\beta\tilde{F}). \tag{2.16}$$

The population is in equilibrium when $R$ is much larger than $\rho_f$. Define $\delta\tilde{\mathcal{A}}/\beta\delta\tilde{F}$ as the limit of the quantity in the square brackets in Eq. 2.15,

$$\frac{\delta\tilde{\mathcal{A}}}{\beta\delta\tilde{F}} = \lim_{R\to\infty} \frac{\mathrm{cov}(\tilde{\mathcal{A}}, \beta\tilde{F})}{\mathrm{var}(\beta\tilde{F})} \tag{2.17}$$

Given these definitions, the asymptotic theoretical prediction for systematic errors is that,

$$\Delta\mathcal{A} \sim \frac{\rho_f}{R}\frac{\delta\tilde{\mathcal{A}}}{\beta\delta\tilde{F}}, \tag{2.18}$$

for any observable $\mathcal{A}$, except the free energy. For the free energy, the simpler expression holds,

$$\Delta F = \frac{\rho_f}{2\beta R}. \tag{2.19}$$

Note that $\delta\tilde{\mathcal{A}}/\delta\tilde{F}$ can be interpreted as the slope of the regression line through the joint distribution $p_{\mathcal{A}F}$. To see this let $\langle x \mid y \rangle$ be the conditional average of $x$ given $y$. For a general bivariate normal distribution, the conditional average is given by

$$\langle x \mid y \rangle = \mu_x + \frac{\mathrm{cov}(x,y)}{\sigma_y^2}(y - \mu_y), \tag{2.20}$$

from which one sees that $\mathrm{cov}(\tilde{\mathcal{A}}, \tilde{F})/\mathrm{var}(\tilde{F})$ is the slope of the linear dependence of $\tilde{\mathcal{A}}$ on $\tilde{F}$. One should not, however, consider Eq. 2.19 to be a special case of Eq. 2.18 by setting $\delta\tilde{F}/\delta\tilde{F} = 1$ since the free energy error equation has an extra factor of 1/2.

For weighted averages we expect similar results for systematic errors but with $R$ replaced by $MR_0$, where $R_0$ is the size of the individual runs and $M$ the number of runs in the weighted average. The substitutions $R \to MR_0$ in Eqs. 2.18 and 2.19 should become exact for weighted averages as $R_0/\rho_f \to \infty$ but for finite $R_0/\rho_f$, where the joint distribution is not close to a bivariate Guassian, the dependence on $M$ may be more complicated.

### 2.2.2 Statistical errors

The statistical error $\delta\tilde{\mathcal{A}}$ of an observable $\mathcal{A}$ is the square root of the variance of the estimator

$$\delta\tilde{\mathcal{A}} \equiv [\mathrm{var}(\tilde{\mathcal{A}})]^{1/2}. \tag{2.21}$$

Statistical errors scale inversely in the square root of the number of independent observations. In the absence of resampling, the number of independent measurements in PA is the population size $R$. However, the resampling step makes identical copies of replicas and thus correlates the

population so that the effective number of independent measurements is less than $R$. On the other hand, MCMC sweeps at each temperature decorrelate the replicas. Thus, if we consider only the correlating effect of resampling we will obtain an upper bound on the statistical errors.

Family trees can be constructed for each member of the initial population. Call all the descendants of replica $i$ in the initial population a family and let $\mathfrak{n}_i$ be the fraction of the population in family $i$. In a typical PA simulation starting at infinite temperature and ending at a low temperature the great majority of initial replicas have no descendants, $\mathfrak{n}_i = 0$. To obtain an upper bound that ignores the decorrelating effect of the MCMC sweeps, assume that observable $\mathcal{A}$ takes a single value $\tilde{\mathcal{A}}_i$ for every member of family $i$. If the MCMC algorithm applied at each temperature step were completely ineffectual, this would be the case. Given this assumption, the estimator $\tilde{\mathcal{A}}$ for the full simulation is

$$\tilde{\mathcal{A}} = \sum_i \mathfrak{n}_i \tilde{\mathcal{A}}_i. \tag{2.22}$$

Next make the additional approximation, which leads to an even weaker upper bound, that the variance of the value of the observable in each family is $\mathrm{var}(\mathcal{A})$, the full variance of the observable in the thermal ensemble. In particular we are ignoring the possibility that the observable is correlated with the family size. For a given distribution of family sizes, we obtain the variance of the estimator of the observable $\mathrm{var}(\tilde{\mathcal{A}})$,

$$\mathrm{var}(\tilde{\mathcal{A}}) \leq \mathrm{var}(\mathcal{A}) \sum_i \mathfrak{n}_i^2. \tag{2.23}$$

Note that if every family contained one member and there were $R$ families then $\mathfrak{n}_i = 1/R$ from which we would obtain the result for $R$ independent measurements, that $\delta\tilde{\mathcal{A}} = [\mathrm{var}(\mathcal{A})/R]^{1/2}$. More generally, the statistical errors are bounded by the second moment of the family size distribution. Suppose this moment has a limit for large $R$ and define the *mean square family size*, $\rho_t$,

$$\rho_t = \lim_{R\to\infty} R \sum_i \mathfrak{n}_i^2. \tag{2.24}$$

In terms of $\rho_t$, the bound on the statistical error in $\delta\tilde{\mathcal{A}}$ is

$$\delta\tilde{\mathcal{A}} \leq \sqrt{\mathrm{var}(\mathcal{A})\rho_t/R}. \tag{2.25}$$

The quantity $R/\rho_t$ is an effective number of independent measurements.

A second measure of the effective number of families is related to the entropy $S_f$ of the family size distribution

$$S_f = -\sum_i \mathfrak{n}_i \log \mathfrak{n}_i. \tag{2.26}$$

The exponential $e^{S_f}$ is an effective number of families. Suppose $R/e^{S_f}$ has a limit and define the *entropic family size*, $\rho_s$,

$$\rho_s = \lim_{R\to\infty} R/e^{S_f}. \tag{2.27}$$

The quantity $R/\rho_s$ is an alternative measure of the number of independent measurements. If every family is a singleton then $\rho_s = \rho_t = 1$. If the family size distribution is exponential with mean $\mu$ then it is straightforward to show that $\rho_t = 2\mu$ and $\rho_s \approx 1.53\mu$. As we shall see in Sec. 2.4.2 these two measures are always close to one another. All of the characteristic sizes $\rho_f$, $\rho_s$ and $\rho_t$ are defined as limits as $R$ goes to infinity but, in practice, we measure them at a fixed large $R$.

### 2.2.3 Comparison of errors in PA and PT

In the previous two subsections we have seen that systematic and statistical errors in PA both decrease with population size $R$; systematic errors diminish as $1/R$, while statistical errors diminish as $1/\sqrt{R}$. Population annealing is a sequential Monte Carlo method while the great majority of simulation methods in statistical physics are Markov chain Monte Carlo methods. For MCMC methods observables are measured using time averages rather than ensemble averages as is the case for PA and the equivalent quantity to population size is the length of the run, $\mathcal{T}$. Errors are reduced by increasing the running time and are estimated from the autocorrelation functions of observables. Systematic errors in MCMC diminish as $\exp(-\mathcal{T}/\tau_{\mathrm{exp}})$ where $\tau_{\mathrm{exp}}$ is the "exponential autocorrelaton time," while statistical errors in an observable $\mathcal{A}$ diminish as $\sqrt{2\tau_{\mathrm{int}}^{\mathcal{A}}/\mathcal{T}}$ where $\tau_{\mathrm{int}}^{\mathcal{A}}$ is

the "integrated autocorrelation time" for $\mathcal{A}$, see, for example, Ref. [50] for a discussion of integrated and exponential autocorrelation times.

In a loose sense we can equate the equilibration population size $\rho_f$ with the exponential autocorrelation time $\tau_{\exp}$ and either of the family size measures, $\rho_s$ or $\rho_t$, with the integrated autocorrelation time. Naively, it would appear that even if the measures $\tau_{\exp}$ and $\rho_f$ were comparable, a MCMC method would have a considerable advantage over PA because MCMC algorithms converge exponentially in the amount of computational work rather than inversely. On further reflection, one sees that the exponential advantage of MCMC is mostly illusory because of statistical errors, which decrease only as the inverse square root of the amount of computational work for both MCMC and PA. For both types of algorithms, the systematic errors are dwarfed by the statistical errors for simulations of a single system.

However, for disordered systems, it is usually necessary to carry out an additional average over realizations of the disorder. Statistical errors for disorder averaged quantities decrease with the number of disorder realizations, $n$ as $1/\sqrt{n}$. When $n$ is large enough, there could be a regime where statistical errors in disorder averages are smaller than the systematic errors of PA. To investigate this issue more quantitatively consider an observable $\mathcal{A}$ and its disorder average $[\langle \mathcal{A} \rangle]_d$ where $[\ldots]_d$ indicates a disorder average. Using Eq. 2.18 we have the following expression for the systematic error in the disorder average, $\Delta [\langle \mathcal{A} \rangle]_d$,

$$\Delta [\langle \mathcal{A} \rangle]_d \approx \left[ \frac{\rho_f}{R} \frac{\delta \tilde{\mathcal{A}}}{\beta \delta \tilde{F}} \right]_d . \tag{2.28}$$

Let $\delta [\langle \mathcal{A} \rangle]_d$ be the statistical error in $[\langle \mathcal{A} \rangle]_d$ and suppose that the main contribution to this statistical error comes from the variance with respect to disorder in $\langle \mathcal{A} \rangle$ defined by

$$\Sigma^2_{\mathcal{A}} = \left[ \langle \mathcal{A} \rangle^2 \right]_d - [\langle \mathcal{A} \rangle]^2_d . \tag{2.29}$$

Systematic errors are negligible relative to statistical error if $\Delta\left[\langle\mathcal{A}\rangle\right]_d \ll \Sigma_{\mathcal{A}}/\sqrt{n}$ where $n$ is the number of disorder realizations in the sample. Thus systematic errors are negligable if

$$\frac{\sqrt{n}}{\Sigma_{\mathcal{A}}}\left[\frac{\rho_f}{R}\frac{\delta\tilde{\mathcal{A}}}{\beta\delta\tilde{F}}\right]_d \ll 1, \tag{2.30}$$

and for the free energy we have the simpler expression,

$$\frac{\sqrt{n}}{\Sigma_{\mathcal{F}}}\left[\frac{\rho_f}{2\beta R}\right]_d \ll 1, \tag{2.31}$$

In our $L = 10$ simulations of the Edwards-Anderson model, discussed below, $n = 5000$ and $\Sigma_F = 23.9$ at $\beta = 5$. Our equilibration criteria requires that $\rho_s/R \leq 10^{-2}$ and, for most samples, $\rho_s/R \ll 10^{-2}$. As we shall see in Sec. 2.4.2, $\rho_f$ is typically less than a factor of two larger than $\rho_s$. Thus the LHS of Eq. 2.31 for the disorder average of the free energy is less than $10^{-2}$ and we are safely in the regime where statistical errors greatly exceed systematic errors.

## 2.3 Application to the three-dimensional EA model

Sampling low temperature equilibrium states of the $3d$ EA model is computationally very difficult. It is known that finding ground states of the $3d$ EA models is an NP-hard computational problem and it is believed that sampling low temperature equilibrium states is also exponentially hard in the sense that the amount of computational work needed to achieve a fixed accuracy in sampling grows exponentially in the system size, $L$. For MCMC algorithms, this intuition can be made more precise as a statement about the $L$ dependence of autocorrelation times while for PA it is a statement about quantities such as $\rho_f$, $\rho_t$ and $\rho_s$, introduced in Sec. 2.2, which characterize population sizes required for equilibration.

There are large sample-to-sample variations in the difficulty of sampling equilibrium states of the $3d$ EA model. It is known that the distribution of integrated autocorrelation times and

other equilibration measures for PT is approximately lognormal [56, 2, 134]. One of the important question studied in Sec. 2.4 is whether PA and PT both find the same spin-glass instances to be either hard or easy.

There are two reasons why the $3d$ EA model is computationally difficult that can be understood intuitively in terms of the free energy landscape. The first reason is that the free energy landscape is rough for typical instances with several relevant local minima separated by high barriers. Both parallel tempering and population annealing are designed to partially overcome this source of computational hardness though it certainly plays a role [134]. The second reason is related to temperature chaos [32, 124], which is effectively a change in dominance between minima in the free energy landscape as a function of temperature. At high temperatures free energy minima with large entropies dominate while at lower temperatures free energy minima with low energies dominate and finding these rare low energy states is difficult for both PA and PT.

### 2.3.1 Simulation Details

The large datasets used in the present study were obtained in previous studies of the low temperature phase of spin glasses [133, 125], the dynamics of PT [134], and a comparison of PA and PT for finding ground states [126]. These datasets involves roughly $n \approx 5000$ disorder realizations for each of 5 system sizes, $L = 4$, 6, 8, 10, and 12. The same set of disorder realizations were simulated using both PA and PT to allow for a detailed comparison between the two algorithms. In addition, we carried out PA simulations for $n = 1000$, $L = 14$ instances. The parameters of the PA simulations are given in Table 2.1. In our implementation of PA, the annealing schedule has temperatures that are evenly spaced in $\beta$ starting from infinite temperature. In all PA simulations we used $N_S = 10$ Metropolis sweeps per temperature. The number of Metropolis sweeps per simulation, $\mathcal{W}$ is given by $\mathcal{W} = RN_SN_T$ so that $\mathcal{W}$ is a rough measure of the computational work expended per spin in the simulation. For the $L = 14$ runs we used weighted averaging with $M = 10$ independent runs per instance so that here $\mathcal{W} = MRN_SN_T$.

**Table 2.1.** Parameters of the main population annealing simulations [125]. $L$ is the system size, $R$ is the standard number of replicas, $T_0$ is the lowest temperature simulated, $N_T$ is the number of temperatures (evenly spaced in $\beta$) in the annealing schedule, and $\mathcal{W} = RN_TN_S$ is the number of sweeps applied to a single disorder realization. $n$ is the number of disorder realizations and $n_{\rm hard}$ is the number of hard instances requiring more than $R$ replicas to meet the equilibration requirement. For $L = 14$ we used weighted averaging with $M = 10$ independent runs so $\mathcal{W} = MRN_TN_S$ for this case.

| $L$ | $R$ | $T_0$ | $N_T$ | $\mathcal{W}$ | $n$ | $n_{\rm hard}$ |
|---|---|---|---|---|---|---|
| 4 | $5 \times 10^4$ | 0.20 | 101 | $5 \times 10^7$ | 4941 | 0 |
| 6 | $2 \times 10^5$ | 0.20 | 101 | $2 \times 10^8$ | 4959 | 0 |
| 8 | $5 \times 10^5$ | 0.20 | 201 | $10^9$ | 5099 | 5 |
| 10 | $10^6$ | 0.20 | 301 | $3 \times 10^9$ | 4945 | 286 |
| 12 | $10^6$ | 0.333 | 301 | $3 \times 10^9$ | 5000 | 533 |
| 14 | $3 \times 10^6$ | 0.333 | 401 | $1.2 \times 10^{10}$ | 1000 | N/A |

The equilibration criteria that we used is that $R > 100\rho_s$. The population size for each system size is listed in the column labelled $R$ in Table 2.1. This population size satisfies the equilibration criteria for most disorder realizations. However, for the hardest instances, runs were required with larger population sizes. The number of hard instances, $n_{\rm hard}$ is listed in the last column of the table. The PA simulations were carried out using OpenMP implemented on 8 cores where each core works on a different subset of the population. In addition to the simulations described in Table 2.1 we carried out a detailed study of a single $L = 8$ and a single $L = 4$ disorder realization in which we performed a large number of independent runs for various population sizes to check predictions concerning systematic errors.

The parameters of the PT simulations are given in Table 2.2. In the implementation of PT, the highest temperature is $T = 2$ and each PT sweep involves $N_S = 1$ heat bath sweeps per replica. Each simulation involves $2^{b+1}$ PT sweeps, $2^b$ for equilibration and $2^b$ for data collection. The number of heat bath sweeps per simulation and thus the computational work per spin is $\mathcal{W} = 2^{b+1}N_SN_T$. In fact for computing the overlap $q$, twice this number of sweeps were used because two independent simulations are needed to compute $q$ in PT. Additional details of the PT simulations can be found in Ref. [134].

**Table 2.2.** Parameters of the parallel tempering simulations [133, 134]. $L$ is the system size, $2^b$ is the standard number of Monte Carlo sweeps. $T_0$ is the lowest temperature used, $N_T$ is the number of temperatures, and $\mathcal{W} = 2^{b+1} N_T N_S$ is the number of sweeps applied to a single disorder realization. $n$ is the number of disorder realizations.

| $L$ | $b$ | $T_0$ | $N_T$ | $\mathcal{W}$ | $n$ |
|---|---|---|---|---|---|
| 4 | 18 | 0.20 | 16 | $8 \times 10^6$ | 4941 |
| 6 | 24 | 0.20 | 16 | $5 \times 10^8$ | 4959 |
| 8 | 27 | 0.20 | 16 | $4 \times 10^9$ | 5099 |
| 10 | 27 | 0.20 | 16 | $4 \times 10^9$ | 4945 |

### 2.3.2 Measured Quantities

We measured standard spin glass observables and also quantities intrinsic to the PA algorithm. We measured the energy $\tilde{E}_{\mathcal{J}}$, free energy $\tilde{F}_{\mathcal{J}}$, and spin overlap distribution $\tilde{P}_{\mathcal{J}}(q)$, for all disorder realizations. From $\tilde{P}_{\mathcal{J}}(q)$ we obtained its integral near the origin,

$$\tilde{I}_{\mathcal{J}} = \int_{-0.2}^{0.2} \tilde{P}_{\mathcal{J}}(q). \tag{2.32}$$

From $\tilde{I}_{\mathcal{J}}$ for the $n$ instances we obtain the disorder average $I = [\tilde{I}_{\mathcal{J}}]_d$. Unless required to prevent confusion, we henceforth drop the subscript $\mathcal{J}$ indicating a particular instance. Observables are obtained from population averages in contrast to the situation for PT and other MCMC methods where observables are obtained from time averages. Estimators of observables for a single instance are indicated by a tilde.

We estimated the family-based characteristic sizes, $\rho_t$ and $\rho_s$ for each disorder realization. For the $L = 14$ and for the two individual size $L = 4$ and $L = 8$ instances we also measured the equilibration population size $\rho_f$, which requires multiple runs. Comparison data for PT were obtained in previous studies [133, 134]. For the same set of disorder realizations, we have values of $I_{\mathcal{J}}$ and the integrated autocorrelation time for the spin overlap, $\tau^q_{\mathcal{J},\mathrm{int}}$.

### 2.3.3 Spin overlap measurement

The spin overlap is an important quantity in spin glass studies and its integral near the origin, $I_{\mathcal{J}}$, has been extensively studied as a way of distinguishing competing pictures of the low temperature phase of spin glasses. The measurement of the spin overlap distribution $P_{\mathcal{J}}(q)$ would appear to be computationally twice as difficult as other observables because it requires two independent spin configurations. Indeed, in standard implementations of PT, two separate simulations are run simultaneously and spin configurations from each are combined to obtain values of $q$, so the work required to measure $\tilde{P}(q)$ (and also the link overlap distribution) is twice that for observables obtained from a single spin configuration. In PA, however, it is possible to construct $\tilde{P}(q)$ from a single run by taking advantage of the fact that replicas from different families, i. e. descended from different initial replicas, are independent. We use the following method to estimate $\tilde{P}(q)$ at a given temperature $\beta$.

First, a random permutation of the population, $(\pi_1, \pi_2, \ldots, \pi_{\tilde{R}_\beta})$ is constructed and used to make an initial pairing of replicas in the population. A random permutation is likely to include pairs chosen from the same family. If replica $k$ and replica $\pi_k$ are in the same family they are potentially correlated. This 'incest' problem is corrected sequentially by performing transpositions as needed. Suppose $k$ is the least integer such that replicas $k$ and $\pi_k$ are in the same family. Then the successive replicas $\pi_{k+1}, \pi_{k+2} \ldots$ are tested until the first $j$ $(j > k)$ is found such that replica $\pi_j$ is in a different family than replica $k$ and also replica $\pi_k$ is in a different family than replica $j$. The permutation is now modified by transposing $\pi_j$ and $\pi_k$. This process is continued until there are no more incestuous pairs. Each pair then contributes one value to the histogram for $\tilde{P}(q)$. Notice that in each step of the procedure the number of incestuous pairs decreases by one. So long as the maximum family size is less than $\tilde{R}_\beta/2$, which is required anyway for a well-equilibrated run, this procedure will find an unbiased, non-incestuous pairing. Although the worst case complexity of the procedure is $R^2$, in practice the complexity is order $R$.

Weighted averaging may also be used to combine results for $\overline{P}_{\mathcal{J}}(q)$ from many runs with $P_{\mathcal{J}}(q)$ playing the role of the observable $\mathcal{A}$ in Eq. 2.1. The justification for weighted averaging based on unnormalized population annealing holds though the argument also requires the fact that each family in unnormalized PA is independent and identically distributed.

## 2.4 Results

In this section we present results for both population annealing and parallel tempering (PT). This section serves two purposes. The first purpose is to validate population annealing and verify claims made in Sec. 2.2 about its statistical and systematic errors. The second purpose is to compare the efficiency of PA and PT.

### 2.4.1 Spin overlap

Figure 2.1 shows scatter plots for sizes $L = 4$, 6, 8 and 10 of $\tilde{I}_{\mathcal{J}}$ for both algorithms, with the vertical position of each point the value of $\tilde{I}_{\mathcal{J}}$ for PT and the horizontal position the value for PA. Disorder realizations with $\tilde{I}_{\mathcal{J}} = 0$ for either algorithm are not shown. This figure demonstrates reasonable agreement between the two algorithms for each disorder realization. Note that PT is capable of measuring smaller values of $\tilde{I}_{\mathcal{J}}$ than PA because the number of measurements $2^b$ for PT is larger than the number of measurements $R$ for PA.

Next we consider $I = [\tilde{I}_{\mathcal{J}}]_d$, the disorder average of the integral of the spin overlap in the range from $-0.2 < q < 0.2$. Table 2.3 gives results for both PA and PT for $I$. The quoted errors are obtained from the sample standard deviation divided by the square root of the sample size $\sqrt{n}$ so it is not surprising that the difference between the PA and PT results is much less than the error since both algorithms use the same set of disorder realizations. It is comforting that the results are so close. Since both algorithms are quite different and use different criteria for equilibration it suggests that systematic errors are minimal and cannot be detected in disorder averages with a sample size of 5000.

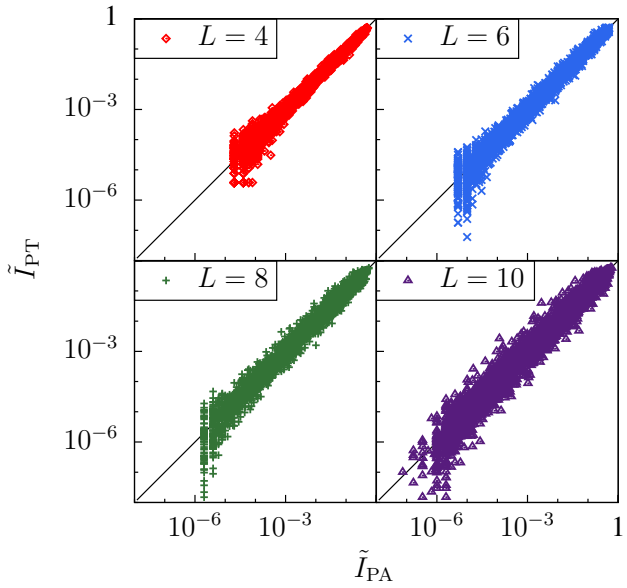


**Figure 2.1.** Log-log scatter plots of $\tilde{I}_{\mathcal{J}}$. Each point represents a disorder realization. The horizontal position of the point is $\tilde{I}_{\mathcal{J}}$ measured in PA and the vertical position is the value of $\tilde{I}_{\mathcal{J}}$ measured in PT, for sizes, $L = 4$, 6, 8 and 10 at $T = 0.2$.

**Table 2.3.** Comparison of the disorder averaged overlap weight near the origin, $I$ between PA and PT at $T = 0.2$ for the same set of disorder realizations.

| $L$ | 4 | 6 | 8 | 10 |
|---|---|---|---|---|
| PA | 0.0186(10) | 0.0194(10) | 0.0205(10) | 0.0200(10) |
| PT | 0.0185(9) | 0.0196(9) | 0.0205(10) | 0.0198(10) |

### 2.4.2 Characteristic population sizes in PA and correlation times in PT

Next we consider quantities that are intrinsic to each algorithm and that are related to errors. Figure 2.2 is a logarithmic scatter plot of $\rho_s$ the entropic family size measured in PA and $\tau^q_{\mathrm{int}}$, the integrated autocorrelation time for the spin overlap measured in PT. Each point represents a disorder realization; the horizontal position of the point is $\log_{10} \rho_s$ and the vertical position is $\log_{10} \tau^q_{\mathrm{int}}$. It is striking that these two quantities are strongly correlated. Both $\rho_s$ and $\tau^q_{\mathrm{int}}$ are related to statistical errors in their respective algorithms and large values correspond to hard instances that require lots of computer resources to simulate accurately. It is clear that the hardness of an instance for PA and for PT are strongly correlated.

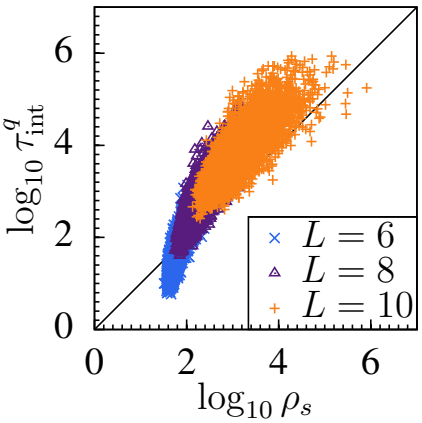


**Figure 2.2.** Log-log scatter plot of $\rho_s$, entropic family size for PA vs. $\tau^q_{\rm int}$ the integrated autocorrelation time of the spin overlap for PT. Each point represents a single disorder realization and there are roughly 5000 disorder realizations each for sizes $L = 6$, 8 and 10 at $T = 0.2$.

Figure 2.3 shows histograms of $\log_{10}\rho_s$ (left panel) and $\log_{10}\tau^q_{\rm int}$ (right panel) for all 4945 disorder realizations of size $L = 10$ at $T = 0.2$. Both distributions are very broad and both are skewed toward hard disorder realizations though the $\rho_s$ distribution is more sharply peaked than the $\tau^q_{\rm int}$ distribution. Figure 2.4 is a log-linear plot of the disorder averages $[\log_{10}\rho_s]_d$ and

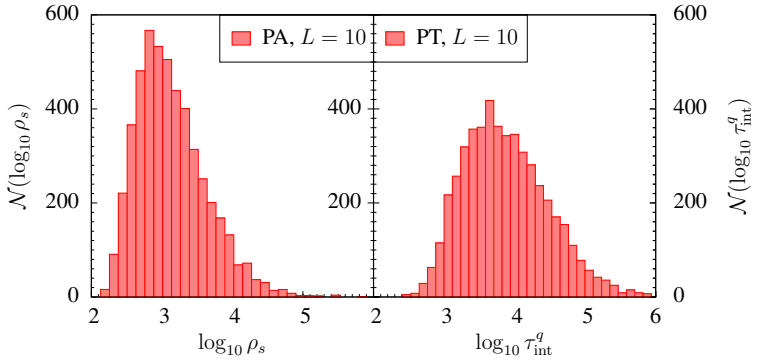


**Figure 2.3.** Histogram of $\log_{10}\rho_s$ (left panel) and $\log_{10}\tau^q_{\rm int}$ (right panel) for all 4945 disorder realizations, size $L = 10$ at $T = 0.2$.

$[\log_{10}\tau_{\rm int}]_d$ vs. system size $L$. Square symbols (blue) are for PT at $T = 0.2$, circular symbols (red) for PA at $T = 0.2$ and triangular symbols (green) for PA at $T = 0.42$. The nearly linear behavior

suggests that both algorithms suffer exponential slowing with system size as expected. The fitted slope is greater for PT than for PA, however, one should be cautious in interpreting these fits as indicating better scaling for PA relative to PT. There is some upward curvature for PA in the data for both temperatures so the asymptotic scaling slope may be significantly larger than the finite $L$ slope. In addition, $\tau_{\rm int}$ and $\rho_s$ are not strictly comparable quantities and, finally, neither algorithm has been carefully optimized. Nonetheless, one can safely conclude that PA is at least comparable in efficiency to PT for the sizes studied.

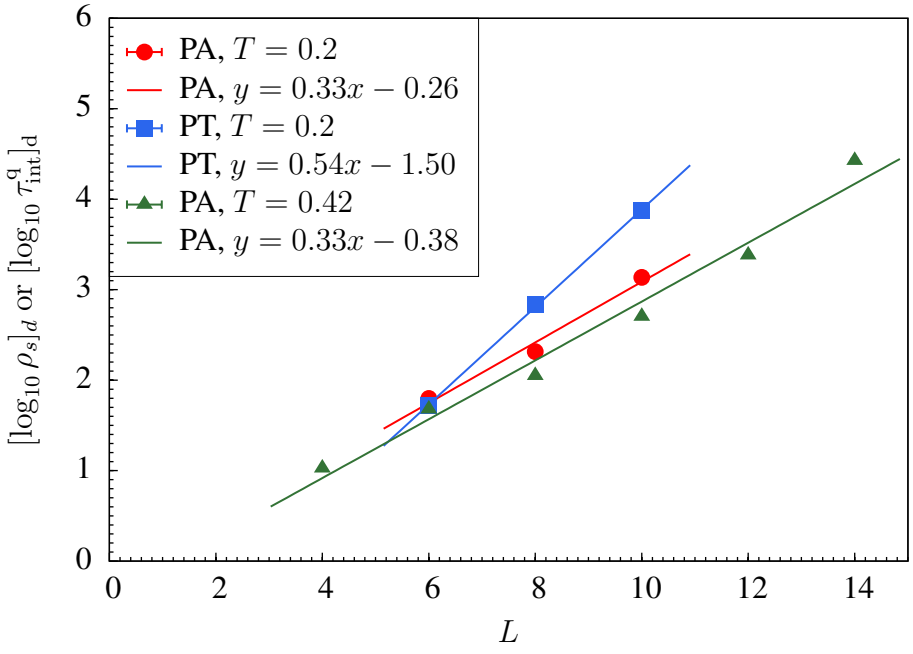


**Figure 2.4.** Disorder averages, $[\log_{10}\rho_s]_d$ for PA and $[\log_{10}\tau^q_{\rm int}]_d$ vs. $L$. Square symbols (blue) are for PT at $T = 0.2$, circular symbols (red) for PA at $T = 0.2$ and triangular symbols (green) for PA at $T = 0.42$. Straight lines are best linear fits to the data.

In Secs. 2.2.1 and 2.2.2 we introduced three characteristic population sizes, $\rho_s$, $\rho_t$ and $\rho_f$. Both $\rho_s$ (see Eq. 2.27) and $\rho_t$ (see Eq. 2.24) are obtained from the distribution of family sizes and are related to statistical errors while $\rho_f$ (see Eq. 2.16) is obtained from the variance of the free energy estimator and controls systematic errors. What is the relation between these three quantities for spin glasses? Figure 2.5 is a scatter plot of $\rho_s$ vs. $\rho_t$ for system sizes $L = 4$, 6, 8 and 10. Each point

represents a single disorder realization. It is clear that these two measures are strongly correlated with $\rho_s$ serving as a lower bound for $\rho_t$.

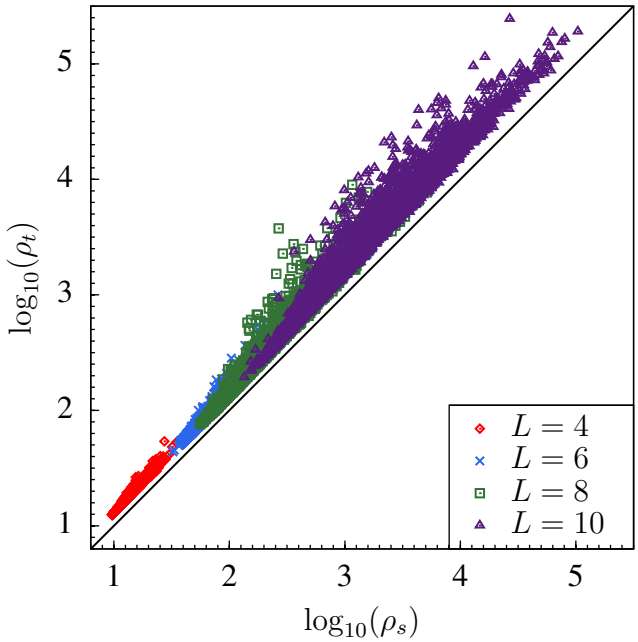


**Figure 2.5.** Scatter plot of $\rho_s$, entropic family size vs. $\rho_t$ the mean square family size for sizes $L = 4$, 6, 8 and 10 at $T = 0.2$.

Figure 2.6 is a scatter plot of $\rho_s$ vs. $\rho_f$ for the $n = 1000$ disorder realizations of size $L = 14$ where each point represents a disorder realization. The value of $\rho_f$ is estimated as $R$ times the sample variance of $\beta\tilde{F}$ from the 10 runs. Since it is obtained from only 10 runs, $\rho_f$ has large statistical errors. The straight line is a best fit through the data points. It is clear that $\rho_s$ and $\rho_f$ are strongly correlated though $\rho_f$ is on average a factor of 1.6 larger than $\rho_s$. The strong correlation between $\rho_s$ and $\rho_f$ justifies using $R > 100\rho_s$ as an equilibration criteria. In principle, equilibration (systematic error) is controlled by $\rho_f$ but measuring $\rho_f$ requires multiple runs whereas measuring $\rho_s$ requires only a single run. Thus, except for situations where weighted averaging is used, it is more straightforward and reasonably well-justified to require that the population size is some factor larger than $\rho_s$. Because $\rho_t$ is just as easy to measure as $\rho_s$, and $\rho_t > \rho_s$, and $\rho_t$ is more

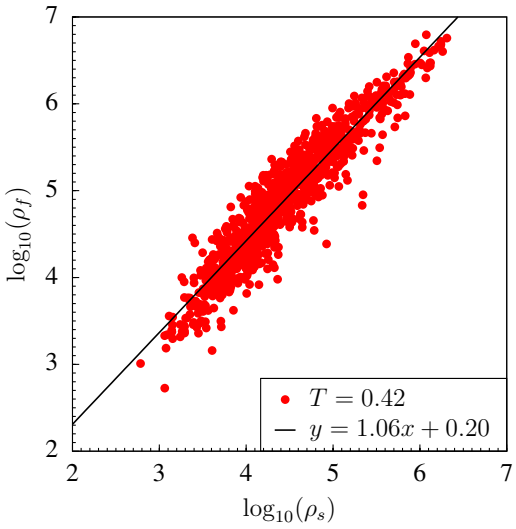


**Figure 2.6.** A scatter plot of the entropic family size, $\rho_s$ vs. equilibration population size $\rho_f$ for 1000 disorder realizations of size $L = 14$ at $T = 0.42$. The straight line is a best fit through the data.

directly related to statistical errors, it may preferable to use $\rho_t$ rather than $\rho_s$ as an equilibration criterion in future simulations.

### 2.4.3 Convergence to equilibrium

Since statistical errors are much larger than systematic errors, in order to investigate systematic errors, i. e. convergence to equilibrium, it is necessary to carry out a very large number of independent simulations of the same disorder realization. From these many runs, systematic errors can be studied as a function of population size $R$. In this section we examine in detail the convergence to equilibrium for two disorder realizations. One of these disorder realizations is the hardest $L = 8$ sample as measured by $\rho_s$. This disorder instance was also studied in detail in Chapters 3 and 4. We call this disorder realization "instance **J8**." The second is an $L = 4$ disorder realization that we call "instance **J4**." Observables of the two instances are shown in Table 2.4.

We first carefully examine, for instance **J8**, the convergence to equilibrium as a function of $R$ for the energy estimator $\tilde{E}$ and the dimensionless free energy estimator $\beta\tilde{F}$ at temperature $T_0 = 0.2$.

**Table 2.4.** Equilibrium values of observables at $T = 0.2$ for the two disorder instances studied in detail, **J4** and **J8**, of sizes 4 and 8, respectively.

| | $\rho_f$ | $-\beta F$ | $-E$ | $\delta\tilde{E}/\beta\delta\tilde{F}$ | $I$ | $\delta\tilde{I}/\beta\delta\tilde{F}$ |
|---|---|---|---|---|---|---|
| **J4** | 33 | 584.138 | 116.541 | 0.0542 | 0.0929 | 0.0843 |
| **J8** | $9.0 \times 10^3$ | 4457.53 | 890.186 | 0.0355 | 0.00104 | 0.00105 |

Figure 2.7 shows histograms of the deviation of the dimensionless free energy estimator from its equilibrium value, $\Delta\beta\tilde{F}$ (top row), the deviation of the energy estimator from its equilibrium value, $\Delta\tilde{E}$ (middle row), and a scatter plot of their joint distribution (bottom row) for population sizes, $R = 10^3$, $10^4$, $10^5$ and $10^6$ (from left to right). For each population size we carried out $\mathcal{M} = 1000$ independent simulations of **J8**. The 'exact' equilibrium values, listed in Table 2.4, are obtained from a weighted average of the 1000 runs at the largest population size, $R = 10^6$. For the two smaller populations the distributions are highly non-Gaussian but as $R$ increases the joint distribution approaches a bivariate Gaussian distribution. The joint distribution initially consists of two well-separated peaks representing the fact that for small $R$ most or all of the population is frequently stuck in a metastable state with both a higher free energy and higher energy. This bimodal distribution is a feature of this particular disorder realization and explains, in part, the computational hardness of this sample. Since $\rho_f \approx 10^4$, the $R = 10^3$ populations are not equilibrated and the $R = 10^4$ populations are barely equilibrated. Finally, for $R = 10^6$, the populations are reasonably well equilibrated so that the $\tilde{E}$ and $\beta\tilde{F}$ distributions are close to Gaussian and the joint distribution is close to a bivariate Gaussian. The slope of the regression line through the scatter plot representing the $R = 10^6$ joint distribution is an estimator of the quantity $\delta\tilde{E}/\beta\delta\tilde{F}$, which controls the error in the energy estimator, see Eq. 2.18.

We can assess more quantitatively whether $\beta\tilde{F}$ and $\tilde{E}$ are described by a bivariate normal distribution. From the $\mathcal{M} = 1000$ runs, we measured the skewness and kurtosis of both variables. For instance **J8** and $R = 10^6$, the skewness and (excess) kurtosis of the dimensionless free energy runs is 0.047 and 0.043, respectively. Both values are statistically indistinguishable from values that would be obtained from a sample of 1000 normal random variates. The corresponding values of

skewness and kurtosis for the energy are 0.121 and 0.152, respectively. Though larger, both values are consistent with a sample of 1000 normal random variates. The joint distribution is, however, only marginally consistent with a bivariate Gaussian, as measured by the Mardia combined skewness and kurtosis test ($p = 0.06$). For instance **J8** at $R = 10^6$, $R/\rho_f \approx 10^2$. For instance **J4** at population size $R = 10^6$ we have $R/\rho_f \approx 3 \times 10^4$ and from $\mathcal{M} = 5000$ runs the joint distribution cannot be distinguished from a bivariate Gaussian by the Mardia combined test ($p = 0.4$).

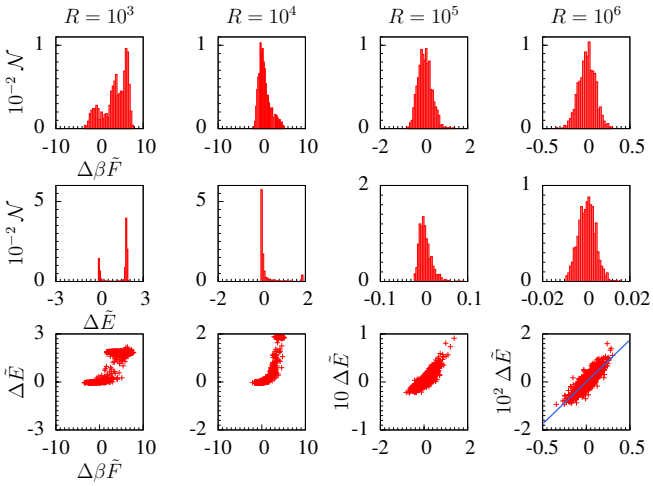


**Figure 2.7.** The top row is histograms of $\Delta\beta\tilde{F}$, the middle row is histograms of $\Delta\tilde{E}$ and the bottom row is scatter plots representing the joint distributions of $\Delta\tilde{E}$ and $\Delta\beta\tilde{F}$ for instance **J8** at $T = 0.2$. Each column is a population size and, from left to right, $R = 10^3$, $10^4$, $10^5$, and $10^6$. The slope of the regression line in the $\Delta\tilde{E}$ vs. $\Delta\beta\tilde{F}$ scatter plot for $R = 10^6$ (lower right box) is the estimator of $\delta\tilde{E}/\beta\delta\tilde{F}$.

Next we study the convergence of the mean values of observables to their equilibrium values as a function of $R$. For each observable $\mathcal{A}$ we obtain the mean value for a single run $\langle\tilde{\mathcal{A}}\rangle$ from a simple average over all $\mathcal{M}$ runs for each population size and we obtain the equilibrium value from a weighted average over all runs at the largest size. Figure 2.8 shows $\langle\Delta\beta\tilde{F}\rangle$, $\langle\Delta\tilde{E}\rangle$, and $\langle\Delta\tilde{I}\rangle$, the deviation of the estimators of the dimensionless free energy, energy and overlap near the origin from their respective equilibrium values, as a function of population size $R$ for instance **J4**. The straight lines are theoretical curves from Eqs. 2.18 and 2.19 using the values of $\rho_f$ and

$\delta\tilde{\mathcal{A}}/\beta\delta\tilde{F}$ estimated at $R = 10^6$ and given in Table 2.4. We see that there is reasonable quantitative agreement with the predicted $1/R$ dependence of systematic errors. The $R = 10^6$ data point is not shown because statistical errors in measuring the exact values $\langle\mathcal{A}\rangle$ are comparable here to systematic errors in $\langle\tilde{\mathcal{A}}\rangle$. Probing the $1/R$ regime of systematic errors proved quite difficult because of the much larger statistical errors. For example, to sufficiently reduce statistical errors for instance **J4** we used $\mathcal{M} = 32000$ independent runs to obtain the $R = 10^5$ averages $\langle\tilde{\mathcal{A}}\rangle$ in Fig. 2.8 and $M = 5000$ independent runs at $R_0 = 10^6$ to obtain 'exact' equilibrium values $\langle\mathcal{A}\rangle$ from weighted averaging.

Figure 2.9 shows similar results for the size 8 instance **J8**. Since the joint distributions are far from bivariate Gaussians for the smaller values of $R$ for instance **J8**, the theoretical predictions for $\langle\Delta\tilde{E}\rangle$, and $\langle\Delta\tilde{I}\rangle$ are poor for the smaller population sizes. The points for $R = 10^6$ in Fig. 2.9 are in essentially perfect agreement with the theoretical predictions of Eq. 2.18, however, since $\rho_f$ and $\delta\tilde{\mathcal{A}}/\beta\delta\tilde{F}$ are all measured at $R = 10^6$ this agreement is really just a check that the joint distribution is close to the assumed bi-variate Gaussian form.

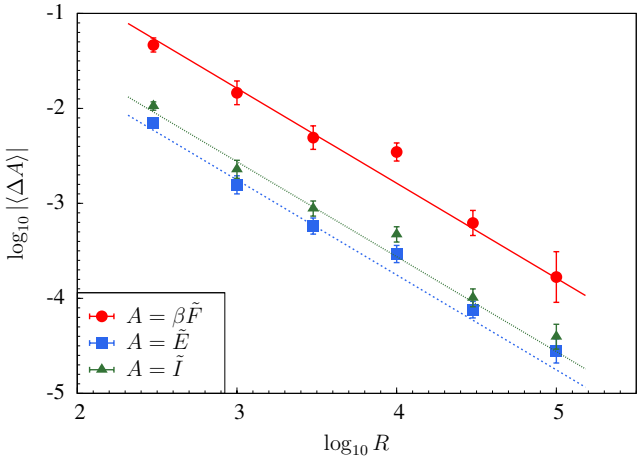


**Figure 2.8.** Log-log plot showing the deviations from equilibrium (systematic errors) in the dimensionless free energy, $\langle\Delta\beta\tilde{F}\rangle$ (red circles), energy, $\langle\Delta\tilde{E}\rangle$ (blue squares) and overlap near the origin $\langle\Delta\tilde{I}\rangle$ (green triangles) as a function of population size $R$ for instance **J4** at $T = 0.2$. The straight lines are theoretical curves based on Eq. 2.18 and 2.19.

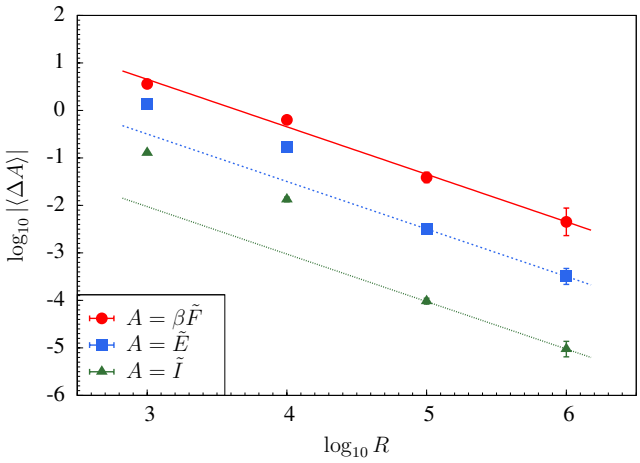


**Figure 2.9.** Log-log plot showing the deviations from equilibrium (systematic errors) in the dimensionless free energy, $\langle\Delta\beta\tilde{F}\rangle$ (red circles), energy, $\langle\Delta\tilde{E}\rangle$ (blue squares) and overlap near the origin $\langle\Delta\tilde{I}\rangle$ (green triangles) as a function of population size $R$ for instance **J8** at $T = 0.2$. The straight lines are theoretical curves based on Eq. 2.18 and 2.19.

Next, we examine the convergence of the various characteristic population sizes to their asymptotic values. Figure 2.10 shows the finite size estimators of $\rho_f$, $\rho_s$ and $\rho_t$ vs. the population size $R$ at which they are measured for instance **J8**. For this instance all of these quantities have values near $10^4$ and their values are near their asymptotic values for the two largest population sizes for which $R \geq 10\rho$. The rapid convergence of $\rho_f$ supports the hypothesis that equilibrium is approached as $1/R$. We do not show a similar graph for instance **J4** since all three $\rho$ measures are already saturated to their asymptotic values within statistical errors even for smallest population sizes studied.

Finally, we can also gain some insights into weighted averaging from the detailed study of a single instance. The question we address is whether a single run is significantly better than a weighted average with the same total population size. We computed the systematic error in the weighted average of the dimensionless free energy $\beta\tilde{F}$ of instance **J8** for $R_0 = 10^3$ and $M = 10$ and compared it to the systematic error for a single run with $R = MR_0 = 10^4$. We used $\mathcal{M} = 1000$ independent runs with $R = 10^3$ to compute the mean $\langle\overline{\beta F}\rangle$ and standard error of the weighted

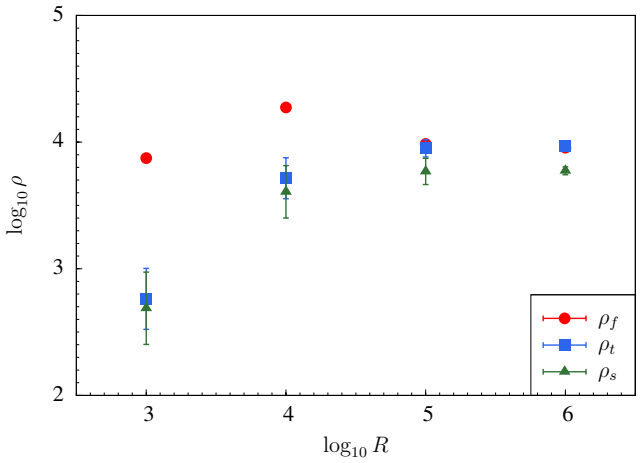


**Figure 2.10.** Log-log plot showing estimators of the equilibration sizes $\rho_f$ (red circles), $\rho_t$ (blue squares) and $\rho_s$ (green triangles) as a function of population size $R$ at $T = 0.2$ for instance **J8**.

average. To compute the mean, we take $M = 10$ random values from the set of $\mathcal{M} = 1000$ runs, compute the weighted average and then take the mean of that weighted average over many such experiments. We used the blocking method to compute the standard error of the mean. We used $\mathcal{M} = 1000$ runs with $R = 10^4$ to obtain to obtain $\langle\beta\tilde{F}\rangle$ and its error. We found that $\langle\Delta\beta\overline{F}\rangle = 0.75 \pm 0.12$, while $\langle\Delta\beta\tilde{F}\rangle = 0.63 \pm 0.04$. Thus, the weighted average has roughly the same systematic errors as the single long run. Note that in this example, $R_0 < \rho_f$. We expect the differences between weighted averaging and a single long run to vanish as $R_0/\rho_f \to \infty$. Unfortunately, even with 1000 independent runs, we did not achieve sufficient statistical power to distinguish the weighted average clearly from a single long run though we expect the former to have somewhat larger systematic errors. These considerations lead to the following conjecture: Suppose one has available a fixed total amount of work defined by a total population size, $R_0 M$ such that $R_0 \gtrsim \rho_f$, then the weighted average obtained from $M$ runs each with population size $R_0$ is statistically indistinguishable from a single long run with population $R = R_0 M$. However, the discussion in Sec. 2.2.3 comparing PA and PT for disorder averaging is relevant here as well. If a sufficiently large disorder sample is simulated, the differences in systematic errors between a

weighted average and a single long run could become relevant. While additional work to understand the systematic errors in weighted averaging is needed, it seems clear that weighted averaging is a useful tool for studying hard problems requiring very large total populations.

## 2.5 Discussion

We have shown that population annealing is an effective and efficient algorithm for simulating spin glasses. It is comparably efficient to parallel tempering, the standard in the field, and it has several advantages. The first advantage is that it is naturally a massively parallel algorithm. The convergence to equilibrium occurs as the population size grows and each replica in the population can be simulated independently. Since realistic spin glass simulations using population annealing require population sizes of the order of $10^6$ or more, there is a much greater scope for parallelism than in parallel tempering where less than 100 replicas are simulated in parallel. To put this difference in perspective, recall that parallel tempering is a Markov chain Monte Carlo (MCMC) algorithm while population annealing is a sequential Monte Carlo (SMC) algorithm. From a computational complexity perspective, when going from a MCMC algorithm to a SMC algorithm, time is exchanged for hardware so that long running times can be exchanged for massive parallelism. The downside of exchanging time for hardware is that population annealing has much larger memory requirements than parallel tempering.

A second advantage of population annealing is access to weighted averaging, which allows multiple independent runs of PA to be combined to improve both statistical *and* systematic errors. Weighted averaging opens the door to distributed computing. It is potentially possible to quickly simulate very difficult to equilibrate instances of spin glasses or other hard statistical mechanical models by distributing the work over a large and inhomogeneous set of computational resources. The only information that needs to be collected and analyzed centrally from each run is the estimators of observables together with the estimator of the free energy.

Apart from its large memory usage, the main disadvantage of population annealing (and SMC methods generally) is that it converges to equilibrium inversely in population size wheres parallel tempering (and MCMC methods generally) converges exponentially. In most situations this difference is moot because statistical errors are much larger than systematic errors. However, for very high precision disorder averages, it is possible that the exponential convergence of parallel tempering would be an advantage over the power law convergence of population annealing.

Thus far the implementations of population annealing for large scale simulations have used a simple annealing schedule. The temperature set is uniform in inverse temperature and there are a constant number of Metropolis sweeps at each temperature. It is plausible that a more complicated annealing schedule might be more efficient. It is perhaps possible that the annealing schedule can be adaptively adjusted to the particular problem instance in analogy to related proposals for parallel tempering simulations [59, 12]. It might also improve efficiency to change the population size with temperature. In addition, our implementation uses the Metropolis algorithm at every temperature, however, at low temperatures kinetic Monte Carlo might be preferable and, at intermediate temperatures cluster moves, might be useful [136].

Population annealing is a general method suitable for simulating equilibrium states of systems with rough free energy landscapes. It can be applied to any system for which there is a parameter, such as temperature, that takes the equilibrium distribution from an easy to simulate region, e.g. at high temperature, to a hard to simulate region, e.g. at low temperature. In addition to spin systems, population annealing may prove useful in simulating the equilibrium states of dense fluids or complex biomolecules.

# CHAPTER 3

# FINDING GROUND STATES OF SPIN GLASSES

Population annealing is a Monte Carlo algorithm that marries features from simulated annealing and parallel tempering Monte Carlo. As such, it is an ideal method to overcome large energy barriers in the free-energy landscape while minimizing a Hamiltonian. Thus, population annealing Monte Carlo can be used as a heuristic to solve combinatorial optimization problems. We illustrate the capabilities of population annealing Monte Carlo by computing ground states of the three-dimensional Ising spin glass with Gaussian disorder, whilst comparing to simulated annealing and parallel tempering Monte Carlo in this chapter. Our results suggest that population annealing Monte Carlo is significantly more efficient than simulated annealing but comparable to parallel tempering Monte Carlo for finding spin-glass ground states. This chapter is adapted from Ref. [126].

## 3.1 Introduction

Spin glasses present one of the most difficult challenges in statistical physics [113]. Finding spin-glass ground states is important in statistical physics because some properties of the low-temperature spin-glass phase can be understood by studying ground states. For example, ground-state energies in different boundary conditions have been used to compute the spin stiffness exponent [24, 43, 47]. More generally, the problem of finding ground states of Ising spin glasses in three and more dimensions is an NP-hard combinatorial optimization problem [4] and is thus closely related to other hard combinatorial optimization problems [80], such as protein folding

[120] or the traveling salesman problem. As such, developing efficient algorithms to find the ground state of a spin-glass Hamiltonian represents an important problem across multiple disciplines.

Many *generally applicable* computational methods have been developed to solve hard combinatorial optimization problems. Exact algorithms that efficiently explore the tree of system states include branch-and-cut [111] algorithms. Heuristic methods include genetic algorithms [41, 42], particle swarm optimization [5] and extremal optimization [14, 77]. The focus of this chapter is on heuristic thermal Monte Carlo methods. In particular, we study simulated annealing [62], parallel tempering [114, 40, 49] and population annealing [48]. The first two methods are well-known and have been successfully applied to minimize Hamiltonians, while the third has been much less widely used in statistical physics and a primary purpose of this work is to introduce population annealing as an effective method for finding ground states of frustrated disordered spin systems.

It is well known that parallel tempering is more efficient at finding spin-glass ground states than simulated annealing [103, 81] because parallel tempering is more efficient at overcoming free-energy barriers. Here we find that population annealing is comparably efficient to parallel tempering and, thus, also more efficient than simulated annealing. Nonetheless, because of the strong similarities between population annealing and simulated annealing, a detailed comparison of the two algorithms is informative and sheds light on the importance of staying near equilibrium, even for heuristics designed algorithms to find ground states.

The outline of this chapter is as follows. We first describe our measured quantities in Sec. 3.2. We then study the properties of population annealing for finding ground states of the Edwards-Anderson model and compare population annealing with simulated annealing in Section 3.3. We conclude by comparing the efficiency of population annealing and parallel tempering in Section 3.4 and present the conclusions in Section 3.5.

## 3.2 Measured quantities

In our SA simulations in this chapter, we use the same simulation parameters of PA with the resampling step turned off. The optimization issues regarding the implementation of SA will be discussed later. For both PA and SA the ground state is presumed to be the lowest energy spin configuration encountered at the end of the simulation.

For SA it is most efficient to do *multiple* runs rather than do one very long run. Thus, the SA results are typically stated as a function of the number of runs $R$. Population annealing is inherently parallel and we report results for a *single* run with a population size $R$. Indeed, choosing the minimum energy configuration among $R$ runs of SA is equivalent to running PA with the same population size but with the resampling step turned off, which justifies using the same symbol $R$ to describe population size in PA and number of runs in SA.

While population annealing is primarily designed to sample from the Gibbs distribution at nonzero temperatures, here we are interested in its performance for finding ground states. We test the hypothesis that the resampling step in PA improves ground state searches as compared to SA. The motivation for this hypothesis is that the resampling step removes high-energy spin configurations and replaces them with low-energy configurations, thus potentially increasing the probability of finding the ground state for a given value of $R$.

To compare PA and SA we investigated the following quantities. For PA let $g(R)$ be the fraction of the population in the ground state for a run with population size $R$. It is understood that $g$ is measured at the lowest simulated temperature. Clearly, the quantity $g(1)$ is simply the probability of finding the ground state in a single run of SA. Let $\mathcal{P}(R)$ be the probability of finding the ground state in a run with population size $R$. For SA, $\mathcal{P}_{\mathrm{SA}}(R)$ is the probability of finding the ground state in $R$ independent runs, i.e.,

$$\mathcal{P}_{\mathrm{SA}}(R) = 1 - [1 - g(1)]^R. \tag{3.1}$$

However, for PA the resampling step tends to reproduce discoveries of the ground state, we have to estimate $\mathcal{P}_{\rm PA}(R)$ from many independent runs of population size $R$. What we actually measured is $N_0$, the number of occurrences of the ground state in the population from which we obtained $g(R) = N_0/R$ in the case of PA and $g(1)$ in the case of SA.

In the limit of large $R$, PA generates an equilibrium population described by the Gibbs distribution so

$$\lim_{R\to\infty} g(R) = g_0, \tag{3.2}$$

where $g_0$ is the fraction of the ensemble in the ground state,

$$g_0 = \frac{1}{Z(\beta)} 2e^{-\beta E_0} = 2e^{-\beta E_0 + \beta F(\beta)}, \tag{3.3}$$

where $E_0$ is the ground state energy Since PA can measure the Helmholtz free energy, thus we have an independent prediction for the limiting value of $g(R)$.

We considered two disorder averaged quantities as well. The first is the probability of finding the ground state, averaged over disorder samples,

$$\eta = \overline{\mathcal{P}}, \tag{3.4}$$

where the overbar indicates a disorder average. The quantity $\eta$ is the primary measure we use to compare the three algorithms.

The second quantity, $\alpha$ is a disorder-averaged measure of accuracy of finding the ground state energy, i.e.,

$$\alpha = 1 - \overline{(E_{\rm min}/E_0)}, \tag{3.5}$$

where $E_{\rm min}$ is the minimum energy found in the simulation, which might *not* be the true ground state energy.

## 3.3 Comparison between PA and SA

### 3.3.1 Finding ground states with population annealing

To compare population annealing and simulated annealing, we need a collection of samples with known ground-state energies. In Chapter 2, we reported on a simulation of approximately 5000 samples of the 3D EA spin glass for size $L = 4$, 6, 8, and 10 using large population runs of PA. Ground-state energies were obtained from these runs by taking the lowest energy encountered in the population at the lowest temperature, $\beta = 5$, using more than adequate resources. We used the data from this large-scale simulation as the reference ground-state energy for each sample and compared the same set of samples for smaller PA and SA runs. The population size and number of temperatures in the reference data set are shown in Table 3.1.

**Table 3.1.** Simulation parameters of the reference simulations of Ref. [125] from which ground states were obtained. $L$ is the linear system size, $R$ is the population size, $N_T$ is the number of temperatures in the annealing schedule, $\min(N_0)$ is the minimum with respect to samples of the number of replicas in the ground state.

| $L$ | $R$ | $N_T$ | $\min(N_0)$ |
|---|---|---|---|
| 4 | $5\times10^4$ | 101 | 3370 |
| 6 | $2\times10^5$ | 101 | 1333 |
| 8 | $5\times10^5$ | 201 | 172 |
| 10 | $10^6$ | 301 | 2 |

Population annealing, like simulated annealing and parallel tempering, is a heuristic method and it is not guaranteed to find the ground state except in the limit of an infinite population size. Nonetheless, we have confidence that we have found the ground state for all or nearly all samples. For an algorithm like PA that is designed to sample the Gibbs distribution at low temperature, the question of whether the true ground state has been found is closely related to the question of whether equilibration has been achieved at the lowest simulated temperature. The candidate ground state is defined as the minimum energy state in the population at the lowest temperature $\beta$. For an equilibrium ensemble, the fraction of the ensemble in the ground $g_0$ is given by the Gibbs distribution, Eq. (3.3). If the number of copies of the found ground state in the low-temperature

population $N_0$ is large and if the population is in equilibrium, then it is unlikely that the true ground state energy has not been found. Because, if we have not found the true ground state, the number of copies of the true ground state, $Rg_0$ would be expected to be even larger than $N_0$. Thus, if we believe the population is in equilibrium at low temperature and if the candidate ground state is found many times in the low-temperature population, then we have high confidence that the candidate is the true ground state.

Of course, it cannot be guaranteed that the population generated by PA is in equilibrium at low temperature. However, the production runs from which we measured ground-state energies passed a stringent thermalization test. We required a large effective number of independent families based on the family entropy. We required $e^{S_f} \geq 100$; additional runs were done for those samples that did not meet these criteria.

In addition to the equilibration test, we recorded the number of copies of the ground state in the population at the lowest temperature and found that for most samples this number is large. A histogram of $N_0/R = g(R) \approx g_0$ of all samples is given in Fig. 3.1 for each system size $L$. The minimum value of $N_0$ for each system size is shown in Table 3.1. For the small fraction (0.7%) of $L = 10$ samples with $N_0 < 10$ we re-ran PA with a ten-fold larger population, $R = 10^7$. In no case did the ground-state energy change. In addition, for the one sample with $N_0 = 2$ we confirmed the ground state using an exact branch and cut algorithm run on the University of Cologne Spin Glass Server [1]. Based on the strict equilibration criteria and the large number of ground states reported in Table 3.1, we are confident that we have found true ground states for all samples.

As an additional check, we compared the disorder averaged ground state energy per spin against values in the literature using the hybrid genetic algorithm [90] and parallel tempering (PT) [103]. The comparison is shown in Table 3.2 and reveals that all three methods yield the same average energy within statistical errors.

A striking feature of Fig. 3.1 is that the fraction of the ensemble in the ground state $g_0$ decreases rapidly as $L$ increases. Thus, for any temperature-based heuristic, including PA, SA, and PT, it

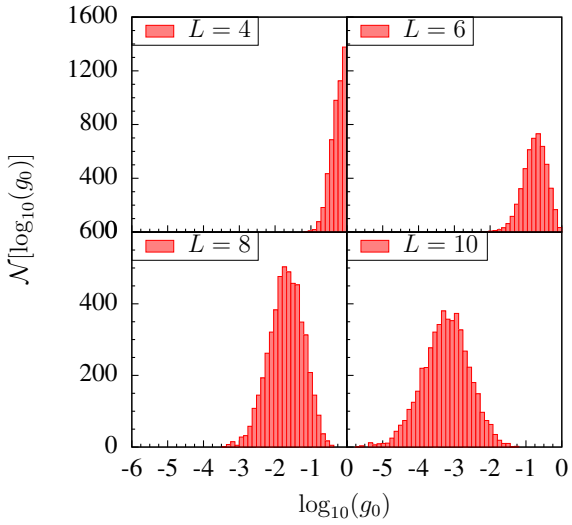


**Figure 3.1.** Histogram of the number of samples with fraction in the ground state $g_0$ at $\beta = 5$ for various sizes $L$, estimated from the reference runs described in Table 3.1. $\mathcal{N}[\log_{10}(g_0)]$ is the number of samples in the logarithmic bin centered at $\log_{10}(g_0)$. There are a total of 50 bins. Note that as $L$ increases, the histograms shift rapidly to smaller values.

**Table 3.2.** Comparison of the disorder averaged ground state energy per spin for the EA model with those obtained from the hybrid genetic algorithm [90] and PT [103].

| $L$ | PA | Hybrid genetic | PT |
|---|---|---|---|
| 4 | -1.6639(14) | -1.6655(6) | -1.6660(2) |
| 6 | -1.6899(7) | -1.6894(5) | -1.6891(4) |
| 8 | -1.6961(5) | -1.6955(4) | -1.6955(6) |
| 10 | -1.6980(3) | -1.6975(5) | -1.6981(7) |

is necessary to simulate at lower temperatures and/or use larger populations (or for PT, longer runs) as $L$ increases. To understand this requirement more formally we re-write Eq. 3.3 in terms of intensive quantities

$$g_0 = 2\exp[-N\beta(e_0 - f(\beta))], \tag{3.6}$$

where $e_0$ and $f(\beta)$ are the ground-state energy and free energy per spin, respectively, and $N = L^3$ is the number of spins. In the thermodynamic limit, $(e_0 - f(\beta))$ is expected to converge to a positive number that is independent of the disorder realization. Thus, for fixed $\beta$, the fraction of the ensemble in the ground state decreases exponentially in the system size.

Population annealing gives a direct estimator of the free energy, thus we can independently measure all of the quantities in Eq. (3.3) and carry out a disorder average. Because the observables on the right-hand side of Eq. (3.3) appear in the exponent, it is convenient to take the logarithm and then carry out the disorder average. Table 3.3 compares $\overline{\log_{10} g_0}$ and $\log_{10} 2 - \beta\overline{(E_0 - F)}/\log(10)$ at $\beta = 5$. The table confirms the expected equilibrium behavior of the fraction in the ground state. Note that the observables $g_0$, $E_0$ and $F$ are not entirely independent quantities, which explains why the statistical errors are significantly larger than the difference in the values. On the other hand, if the simulation was not in thermal equilibrium, these quantities would not agree.

**Table 3.3.** Comparison of the disorder average of the log of the two sides of Eq. (3.3) at $\beta = 5$.

| $L$ | $\overline{\log_{10} g_0}$ | $\log_{10} 2 - \beta\overline{(E_0 - F)}/\log(10)$ |
|---|---|---|
| 4 | -0.2644(28) | -0.2643(28) |
| 6 | -0.7573(46) | -0.7563(46) |
| 8 | -1.6933(77) | -1.6925(67) |
| 10 | -3.2358(104) | -3.2297(91) |

For all the reasons discussed above we believe that we have found the true ground state for all samples. However, our main conclusions would not be affected if a small fraction of the reference ground states are not true ground states.

### 3.3.2 Detailed comparison for a single sample

In this section we present a comparison of population annealing and simulated annealing for a single disorder realization. This sample was chosen to be the hardest sample of $L = 8$ based on having the smallest family entropy, however it has a probability of being in the ground state at the lowest temperature near the average for size $L = 8$. We have confirmed the ground-state energy of this sample found in the reference PA run using the University of Cologne Spin Glass Server [1].

Figure 3.2 shows the fraction of the population in the ground state $g(R)$ as a function of population size $R$ for PA. The result for the probability that SA has found the ground state in a single run is simply the value at $R = 1$. In this simulation, we used $N_T = 101$ temperatures with

$N_S = 10$ sweeps per temperature for both algorithms. It is striking that the fraction of ground states in the population increases by about four orders of magnitude from the small value for SA, $g(1)$ to the limiting value for PA for large $R$, $g(10^6) \approx g_0$. This result shows that resampling greatly increases the probability that a member of the PA population is in the ground state. It suggests that even though equilibration is not required for finding ground states, the probability of finding the ground state is improved when the simulation is maintained near thermal equilibrium. Of course, remaining near equilibrium as the temperature is lowered is also a motivation for SA but lacking the resampling step, SA falls out of equilibrium once the free-energy landscape roughness significantly exceeds $k_{\mathrm{B}}T$. However, the ratio of $g(R)/g(1)$ is an overestimate of the ratio the probabilities for actually finding the ground state for a fixed $R$ because once the ground state is discovered in PA, it is likely to be reproduced many times.

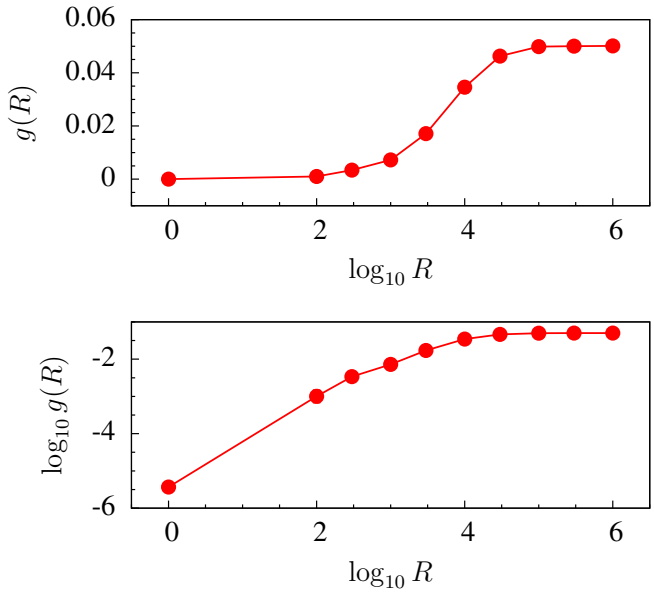


**Figure 3.2.** The fraction of the population in the ground state $g(R)$ as a function of population size $R$ for a single sample using PA with $N_T = 101$ and $N_S = 10$. The point at $\log_{10} R = 0$ corresponds to the probability that a single run of SA will yield the ground state. The upper panel is a log-linear plot and the lower panel is a log-log plot. Error bars are smaller than the symbols.

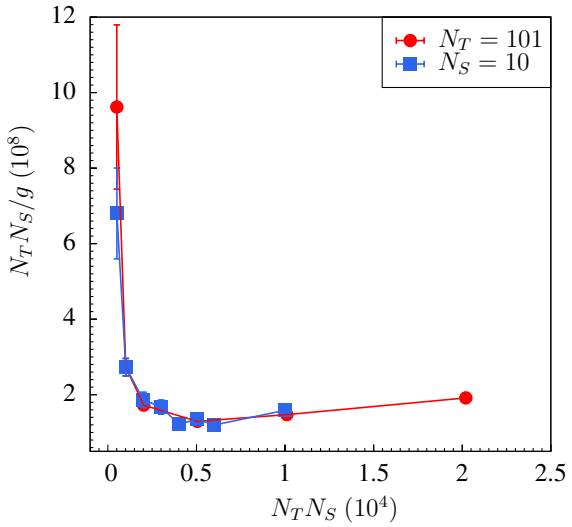


**Figure 3.3.** The computational work divided by the probability of finding the ground state in a single SA run, $N_T N_S/g$ vs the computational work $N_T N_S$ for a single sample. The two curves correspond to holding $N_S = 10$ fixed and varying $N_T$ (blue squares) and holding $N_T = 101$ fixed and varying $N_S$ (red circles). Smaller values of $N_T N_S/g$ correspond to more efficient simulations.

The probability of finding the ground state $\mathcal{P}$ for a given amount of computational work is an appropriate metric to compare the two algorithms. We measured the amount of work $W$ in Metropolis sweeps, $W = RN_T N_S$. In most of our comparisons we used the same value of $N_T$ and $N_S$ for both PA and SA. However, it is not clear whether the two algorithms are optimized with the same chosen values of $N_T$ and $N_S$. We performed additional optimization of SA varying $N_T$ and $N_S$. We used the computational work divided by the probability of finding the ground state in a single SA run, $N_T N_S/g$ as a figure of merit. Note that in the relevant large-$R$ regime, minimizing $N_T N_S/g$ is equivalent to maximizing $\mathcal{P}$ for a fixed amount of work. Figure 3.3 shows $N_T N_S/g$ vs $N_T N_S$ and reveals a broad minimum near $N_T N_S \approx 5 \times 10^3$. We therefore performed SA simulations at the same value used for PA, $N_T N_S = 1010$ and a more nearly optimal value, $N_T N_S = 5000$. Note that for SA it is only the product, $N_T N_S$ that determines the efficiency, not $N_T$ and $N_S$ separately. Note also that the efficiency decreases when $N_T N_S$ is too large suggesting that it is better to do many shorter SA runs rather than a single long run.

Figure 3.4 compares $\mathcal{P}_{SA}$, obtained from Eq. (3.1), and $\mathcal{P}_{PA}(R)$, obtained from multiple runs of PA as a function of the computational work $W$. The upper SA curve corresponds to the optimal value $N_T N_S = 5000$. Computational work was varied by changing $R$ holding $N_T$ and $N_S$ fixed. For intermediate values of $R$, corresponding to realistic simulations, $\mathcal{P}_{PA}$ exceeds $\mathcal{P}_{SA}$ by one or two orders of magnitude and the amount of work needed to be nearly certain of finding the ground state is also more the an order of magnitude less for PA than SA. Note that the effect of optimizing SA is only about a factor of 2. We conclude that for this sample, there is a large difference in the efficiency between PA and SA and this difference cannot be explained by a difference in the optimization of the two methods. To see whether this difference is typical and how it depends on system size, in Sec. 3.3.3 we consider averages over disorder realizations.

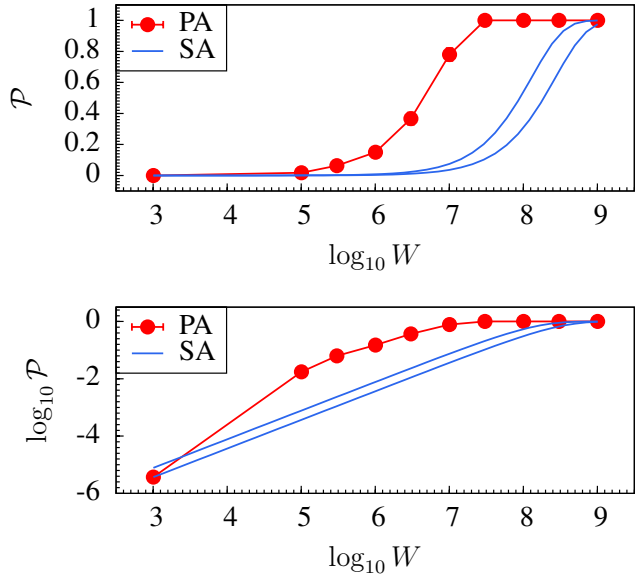


**Figure 3.4.** Probability of finding the ground state $\mathcal{P}$ as a function of the computational work $W = RN_T N_S$ for a single sample for both SA and PA. The computational work is varied by changing population size $R$, holding $N_T N_S$ fixed. For PA and the lower SA curve, $N_T N_S = 1010$ while for the upper SA curve, $N_T N_S = 5000$, which is near the optimum value for SA. The upper panel is a log-linear plot and the lower panel is a log-log plot. Error bars for PA are smaller than the symbols. The SA curves are obtained from Eq. (3.1).

### 3.3.3 Disorder-averaged comparison

We compared population annealing and simulated annealing for approximately 5000 disorder realizations for each of the four system sizes, $L = 4$, 6, 8, and 10, and for several different population sizes. For SA the population size refers to the number of independent runs. Both algorithms use the same annealing schedule with evenly spaced inverse temperatures starting with infinite temperature and ending at $T_0 = 0.2$. The number of sweeps per temperature is $N_S = 10$. The population sizes $R$, number of temperatures in the annealing schedule $N_T$, the number of disorder realizations $n$ and the corresponding parameters for the reference runs are given in Table 3.4.

**Table 3.4.** Parameters of the numerical simulations for comparison between PA and SA. $R$ is the population size, $N_T$ is the number of temperatures, and $n$ is the number of samples studied. The reference parameters are for the PA runs used to estimate the ground state energy for each sample.

| $L$ | $\log_{10} R$ | $N_T$ | $n$ | Ref. $R$ | Ref. $N_T$ |
|---|---|---|---|---|---|
| 4 | {1,2,3,4} | 101 | 4941 | $5\times10^4$ | 101 |
| 6 | {1,2,3,4,5} | 101 | 4959 | $2\times10^5$ | 101 |
| 8 | {1,2,3,4,5} | 101 | 5099 | $5\times10^5$ | 201 |
| 10 | {1,2,3,4,5} | 201 | 4945 | $10^6$ | 301 |

Figure 3.5 shows the disorder averaged error in finding the ground state $\alpha$ [see Eq. (3.5)] as a function of population size $R$ for SA and PA. For small systems neither algorithm makes significant errors even for small populations but as the system size increases, PA is significantly more accurate. Figure 3.6 shows the disorder-averaged fraction of samples for which the ground state is found $\eta$ [see Eq. (3.4)] as a function of population size $R$. Again, we see that for $L = 4$ and 6, the two algorithms are comparable but for $L = 8$ and 10, population annealing is far more likely to find the ground state for the same population size. It is clear from these two figures that population annealing is both more accurate and more efficient at finding ground states than simulated annealing and that as size increases, the relative advantage of PA over SA increases.

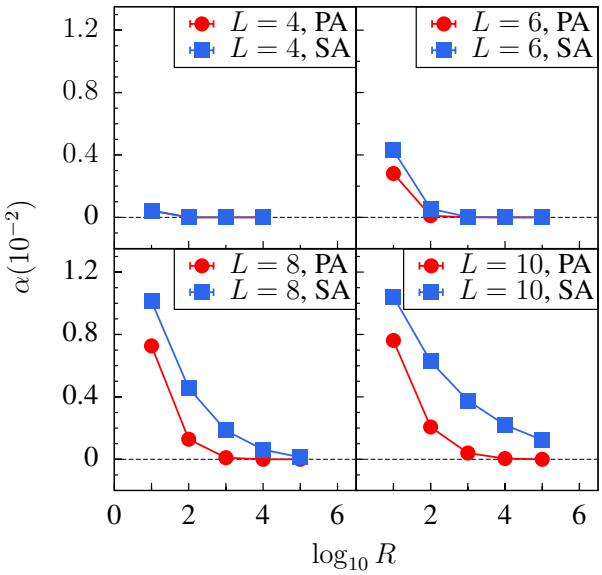


**Figure 3.5.** Error in approximating the the ground state energy, $\alpha$ vs log population size, $\log_{10}(R)$.

## 3.4 Comparison between PA and PT

In this section, we compare the efficiency of population annealing and parallel tempering. Parallel tempering simultaneously simulates $N_T$ replicas of the system at $N_T$ different temperatures. In addition to single temperature Metropolis sweeps, PT uses *replica exchange* moves in which two replicas at neighbouring temperatures swap states. The results for comparison are PT are taken from Ref. [103]. They gave an empirical fit of their data of the form,

$$\eta = \frac{e^{qx}}{1 + e^{qx}}, \tag{3.7}$$

where $q$ is a fitting parameter and $x$ is a function of the computational work $W$ and system size $L$ defined as

$$x = [\log(W/2) - (bL^c - a)]/L^d, \tag{3.8}$$

and the work is calculated in units of Monte Carlo sweeps. For PT, the computational work is given by, $W = N_T N_S$, while for PA, it is given by, $W = RN_T N_S$. We assume that the work

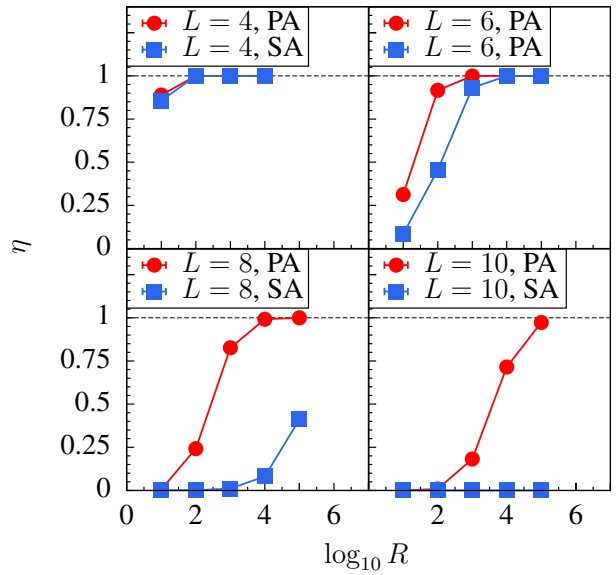


**Figure 3.6.** Fraction of samples for which the ground state is found $\eta$ vs log population size, $\log_{10}(R)$ for population annealing and simulated annealing.

involved in replica exchange for PT and in resampling for PA is negligible compared to the work associated with the Metropolis sweeps. The fitting parameters for the 3d EA model reported in [103] are $a = -0.05$, $b = 1.55$, $c = 1$, $d = 0.2$, and $q = 2$.

Figure 3.7 shows the fraction of samples for which the ground state is correctly found $\eta$ vs the scaled work $x$ for our PA data (points) and the fit for PT (solid curve). It is striking that both algorithms perform nearly identically over the whole range of sizes and amounts of computational work.

## 3.5 Conclusion

We have carried out a detailed comparison of three Monte Carlo heuristics based on thermal annealing for finding ground states of spin-glass Hamiltonians. The algorithms compared are population annealing, simulated annealing and parallel tempering Monte Carlo. We find that population annealing is more efficient than simulated annealing and has better scaling with the system size.

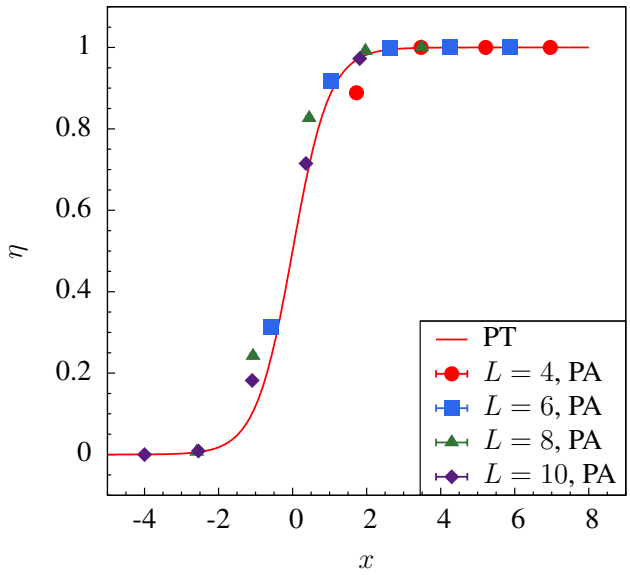


**Figure 3.7.** Fraction of samples for which the ground state is found $\eta$ as a function of the scaled computational work $x$ for both population annealing and parallel tempering. The curve is taken from the empirical fit of Ref. [103].

We also find that population annealing and parallel tempering are comparably efficient. A general conclusion is that Monte Carlo heuristics based on thermal annealing are enhanced by mechanisms that improve thermalization at every temperature. In population annealing this mechanism is resampling and in parallel tempering it is replica exchange. Simulated annealing depends entirely on local Monte Carlo moves and fails to remain close to equilibrium at low temperatures where the free-energy landscape is rough. Furthermore, we observed that the ensemble defined by simulated annealing has far less weight in the ground state than the equilibrium ensemble for realistic computational effort. This deficiency results in a significantly lower probability of finding the ground state for a given amount of computational effort as compared to either population annealing or parallel tempering, which stay close to thermal equilibrium.

There is no obvious reason to suppose that the temperature dependent Gibbs distribution is the best target distribution for improved heuristics such as population annealing or parallel tempering Monte Carlo. Distributions other than the Gibbs distribution that concentrate on the ground state

as “temperature” is decreased might perform even better than the Gibbs distribution and should be investigated.

# CHAPTER 4

# MEASURING FREE ENERGY OF SPIN GLASSES

In this chapter, we study how to measure free energy efficiently using parallel tempering and then we compare the performance with population annealing. An efficient and simple approach of measuring the absolute free energy as a function of temperature for spin-lattice models using a two-stage parallel tempering Monte Carlo and the free energy perturbation method is discussed and the results are compared with those of population annealing Monte Carlo using the three-dimensional Edwards-Anderson Ising spin glass model as benchmark tests. This approach requires little modification of regular parallel tempering Monte Carlo codes with also little overhead. Numerical results show that parallel tempering, even though uses much fewer temperatures than population annealing, can nevertheless equally efficiently measure the absolute free energy by simulating each temperature for longer times. The chapter is adapted from Ref. [123].

## 4.1 Introduction

Measuring free energy is challenging in Monte Carlo simulations, but nevertheless can be very helpful if available. This is not only for the sake of free energy, but also as a consequence entropy can be obtained indirectly since energy can be readily measured in Monte Carlo simulations. More phenomena and finite size effects can be studied or analyzed using free energy and entropy data, including but not limited to the spin stiffness exponents [24, 43, 47], the fractal dimension of droplet excitations in spin glasses [91, 58], the weight of ground states [126], as well as doing weighted averages [70]. A large amount of work has been done for calculating the free energy differences for atomic gases or chemical processes etc. There are both elementary techniques like the free energy

perturbation method [137, 30], thermodynamic integration [109] and the Bennett acceptance ratio method [6], and a wide array of more elaborate techniques including the multicanonical ensemble method [7], umbrella sampling [118], Jarzynski equality [51], Wang-Landau [121] and the weighted histogram analysis method [66] etc.

In this chapter, I will discuss an efficient and simple approach to measure the absolute free energy as a function of temperature for spin-lattice models using a two-stage parallel tempering Monte Carlo and the free energy perturbation method [70]. The method allows one to get access to the absolute free energy with little modification of regular parallel tempering Monte Carlo codes and with also little overhead. The motivation of the work is from our systematic studies of an efficient Monte Carlo algorithm called population annealing Monte Carlo which can measure free energy accurately using the free energy perturbation method [70, 127, 125, 126]. So it is worth exploring whether one can also measure the absolute free energy in parallel tempering simulations using this method. I will discuss more of these two Monte Carlo methods and the challenges in measuring free energy in parallel tempering in the next paragraph. It is however worth mentioning that some of the above listed methods work fundamentally differently from both parallel tempering and population annealing, both of which are thermal Monte Carlo methods. For example, the Wang-Landau algorithm gets access to free energy via a self-consistent sampling of density of states. Since parallel tempering is widely used in spin lattice simulations, I will now only discuss here some of the methods that can be also implemented with parallel tempering, for example the thermodynamic integration [117] and the multiple histogram reweighting [34, 33] methods. Thermodynamic integration has the drawback of integration errors and therefore needs data at many temperatures to be accurate. When data is available at only a few temperatures, interpolation techniques are often needed, making the method less accurate. This is especially a problem when the specific heat shows complicated multi-peak structures like in spin glass [124] and protein folding [119] problems due to various reasons like phase transitions and temperature chaos. The multiple histogram reweighting method is another known effective method and can be

used to compute the free energy difference as well as density of states. This method however is much more complicated than the free energy perturbation method since it requires to record the distribution of energy at many different temperatures and also the autocorrelation times to do the reweighting. There is also a final stage of solving non-linear equation systems. Nevertheless, when density of states is needed, this is a promising technique to use. Since the free energy perturbation method is both simple and effective, I will focus on this method in this work.

For systems with complicated free energy landscapes like spin glasses, a simple Monte Carlo method is not sufficient. Two algorithms that both have a hierarchical structure are shown to be efficient: parallel tempering [114, 40, 49] and population annealing [48, 70]. It has been known for more than a decade that population annealing can accurately estimate the absolute free energy [70] using the free energy perturbation method. But slightly negative view has been held for parallel tempering due to mainly two reasons: the first is that there are too few temperatures, therefore the measurement would not be very accurate while the other is that the highest temperature is usually finite for parallel tempering while in population annealing this can be naturally chosen as infinity. The second problem can be trivially solved by adding a high temperature stage before the parallel tempering stage using simple single temperature Monte Carlo since thermal equilibration is fast at high temperatures. For the first problem, here we show numerically that the larger temperature steps in parallel tempering compared with population annealing are compensated by simulating each temperature for longer times, therefore, permitting equally accurate measurements of free energy in parallel tempering. Parallel tempering however does suffer some minor technical problems, but nevertheless can be readily solved to get access to free energy.

This chapter is organized as follows: Section 4.2 introduces the two-stage free energy measurement algorithm in parallel tempering. The results of parallel tempering and population annealing are compared in Sec. 4.3 and the conclusions are stated in Sec. 4.4.

## 4.2 The two-stage parallel tempering Monte Carlo

We have already discussed the free energy perturbation method for measuring free energy in Sec. 1.3.5. Since parallel tempering (PT) doesn't usually work from $\beta = 0$, we need to modify the parallel tempering algorithm somewhat to get the absolute free energy. One simple way is to implement the simulation in two stages with a high temperature stage using simple single temperature Monte Carlo and a regular parallel tempering Monte Carlo stage. The two stages together will cover from $\beta = 0$ to a low temperature of interest.

- *Simple Monte Carlo stage*

  Suppose the regular parallel tempering Monte Carlo works between $T_{\min}$ and $T_{\max}$, the minimum and maximum temperatures, respectively. The first stage is to use simple single temperature Monte Carlo to simulate the system and work from $\beta = 0$ to $1/T_{\max}$, not including $1/T_{\max}$. Then the free energy can be integrated from $\beta = 0$ to $1/T_{\max}$, including $1/T_{\max}$.

- *Parallel tempering Monte Carlo stage*

  Simulate the system using the regular parallel tempering Monte Carlo between $T_{\min}$ and $T_{\max}$. Then the free energy can be further integrated to the low temperature spin glass phase.

In my implementation, I used $N_\beta$ temperatures evenly distributed in $\beta$ for the first stage and used $N_T$ temperatures evenly distributed in $T$ for the second stage. A uniform distribution in $\beta$ for the first stage is easier to work with because of the infinite temperature. For the second stage, a uniform distribution in $T$ can yield better performance of parallel tempering since the swap probabilities are more constant at different temperatures. Also, no Monte Carlo sweeps were done at $\beta = 0$, it is sufficient to generate random states and make measurements like that of population annealing. In this work, the amount of work is counted in terms of Monte Carlo sweeps and one Monte Carlo sweep is a sequential update of all the spins for one replica at one temperature once. The same number of sweeps were done for both the thermal equilibration run and the data collection run.

It is important to stress here that *the only difference of this algorithm compared with the regular parallel tempering algorithm is the additional simple Monte Carlo stage.* Since the free energy landscape is not rough at high temperatures, the autocorrelation time in sweeps therefore doesn't depend on the energy landscape and system size. It is not necessary to do as many sweeps as in the second stage, where the free energy landscape is rough. Therefore when the algorithm is well optimized, most of the work would be spent in the second stage and the overhead would be small when adding the first stage. In fact, the work spend in the first stage compared with the second stage should vanish in the thermodynamic limit. In addition, due to the similarity of this algorithm with the regular parallel tempering algorithm, it is straightforward to modify a regular parallel tempering code to collect the free energy data.

Finally, I want to point out that there is a technical issue of overflow when computing $Q$ for large systems if $N_T$ is small or there are too many data collection steps for parallel tempering. The later problem can be easily solved since data collection after each Monte Carlo sweep is neither necessary nor desired. Another method is averaging $Q$ on the fly. For the former problem, one can subtract a low energy from the energy of a state when computing the exponential term and then add it back when integrating $-\beta F$. One can also use reasonably more temperatures, but this requires more computational work and may also increase the round trip time. Note that this problem could appear as well for population annealing. However, since population annealing naturally uses a lot more temperatures and the temperatures are usually evenly distributed in $\beta$, no such problem was encountered in our studies of the population annealing algorithm at least up to size $L = 14$ down to $\beta = 3$ [127].

## 4.3 Results

The free energy of each sample was measured using both parallel tempering and population annealing from infinite temperature to low temperatures deep in the spin glass phase. The parallel tempering results are compared with the production run of population annealing [125]. The

**Table 4.1.** Parameters of the reference runs of PA [125] for different system sizes $L$ with periodic boundary conditions. $R$ represents the number of replicas, $1/\beta_0$ is the lowest temperature simulated, $N_T$ is the number of temperatures used in the annealing schedule, $N_S$ is the number of sweeps per temperature and $n$ is the number of samples.

| $L$ | $R$ | $1/\beta_0$ | $N_T$ | $N_S$ | $n$ |
|---|---|---|---|---|---|
| 4 | $5\,10^4$ | 0.2 | 101 | 10 | 100 |
| 6 | $2\,10^5$ | 0.2 | 101 | 10 | 100 |
| 8 | $5\,10^5$ | 0.2 | 201 | 10 | 100 |
| 10 | $10^6$ | 0.2 | 301 | 10 | 100 |

**Table 4.2.** Parameters of PT for different system sizes $L$ with periodic boundary conditions. $N_\beta$ represents the number of temperatures in the simple Monte Carlo stage, $N_T$ is the number of temperatures in the PT stage, $T_{\min}$ is the lowest temperature simulated, $T_{\max}$ is the highest temperature simulated in the PT stage, $N_S$ is the number of sweeps per temperature and $n$ is the number of samples.

| $L$ | $N_\beta$ | $N_T$ | $T_{\min}$ | $T_{\max}$ | $N_S$ | $n$ |
|---|---|---|---|---|---|---|
| 4 | 5 | 10 | 0.2 | 2.0 | $5\,10^5$ | 100 |
| 6 | 5 | 20 | 0.2 | 2.0 | $10^6$ | 100 |
| 8 | 10 | 30 | 0.2 | 2.0 | $5\,10^6$ | 100 |
| 10 | 10 | 40 | 0.2 | 2.0 | $5\,10^7$ | 100 |

reference simulation parameters of population annealing are summarized in Table 4.1 while the simulation parameters of parallel tempering are summarized in Table 4.2. Since the free energy results of the two algorithms agree to a very high degree of precision, I have therefore studied a random subset of samples of [125], 100 samples per system size. This is sufficient to show the equally efficiency of parallel tempering and population annealing in measuring free energy. In the following, I will first make a detailed comparison for a single hard sample and then a large scale comparison for the two algorithms.

### 4.3.1 Detailed comparison of a single hard sample

In this section, I will compare the efficiency of parallel tempering and population annealing in measuring free energy using the hardest sample out of about 5000 samples of $L = 8$ [125]. I will first focus on the dimensionless quantity of $-\beta F$ at a low temperature $T = 0.2$ and study how the

mean and the errorbar of the systematic error of $-\beta F$ change as a function of the amount of work $W$. Then I will study how the relative error of the estimated free energy evolves as a function of temperature at a fixed amount of work. The amount of work is counted as the total number of sweeps performed in the simulation.

The systematic error $\Delta(-\beta F)$ is defined as the differences of the estimated $\beta F$ and the exact $-\beta F$ i.e. $\Delta(-\beta F) = -\beta F + \beta F_{\text{exact}}$. The exact $-\beta F$ is *estimated* using the reference runs of population annealing [125]. The population annealing comparison data with different amount of work is taken from Chapter 3 of finding ground states. The work is varied by changing the number of replicas while holding $N_T = 101$ and $N_S = 10$ constant. The parallel tempering data was done using the parameters of Tabel 4.2 but varying $N_S$. The result is shown in Fig. 4.1. One thing we can learn from this study is that the accuracy is about the same considering neither algorithms is very carefully optimized. Furthermore, both algorithms underestimate the value of $-\beta F$ when the amount of work is too small i.e. the sample is not in thermal equilibrium. This can be understood as those runs are so small so that some low energy states were probably not properly found so that $Q$ was underestimated, and therefore also $-\beta F$.

It is also important and interesting to study how the relative error of the estimated free energy behaves as temperature is lowered due to the integration nature of the method. The relative error is defined as minus the ratio of the standard deviation $\sigma_F$ and mean $\mu_F$ of the estimated free energy. The minus sign is to ensure the defined relative error is positive. The standard deviation and the mean of $F$ can be estimated from the sample standard deviation and the sample mean from multiple independent runs respectively. In this work, 10 independent runs were used to compute the relative error. Figure 4.2 shows this quantity as a function of inverse temperature $\beta$ for both PA and PT for the largest run of Fig. 4.1. Note that there is no trend that the relative error grows as temperature is lowered, showing the effectiveness of both algorithms. In addition, the magnitude of the relative errors is also about the same.

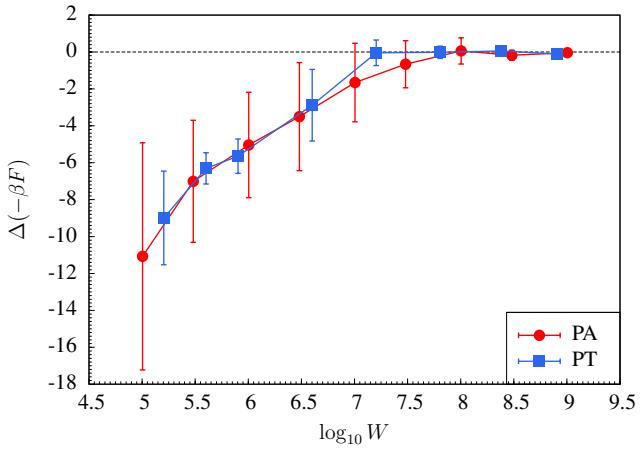


**Figure 4.1.** Log-linear plot of the systematic error $\Delta(-\beta F)$ as a function of the amount of work for a hard sample of $L = 8$ at $T = 0.2$. The errorbar is the the standard deviation of the $-\beta F$ distribution computed from multiple runs, not the errorbar of the sample mean of $-\beta F$. The exact value is taken from a large scale simulation using population annealing [125].

### 4.3.2 A large scale comparison

Before showing the large scale comparison, I would like to show a comparison of the result of $-\beta F$ for a typical sample of each system size in the whole range of temperatures studied. A typical result is shown in Fig. 4.3. Note that the parallel tempering data falls right on top of the population annealing curve, showing the effectiveness of parallel tempering in measuring free energy in a wide range of temperatures.

To make a comparison of more samples, a scatter plot of the free energy per spin $f$ at the lowest simulation temperature $T = 0.2$ of PA and PT is shown in Fig. 4.4. Note that the statistical error compared with the absolute value is too small to be seen in this plot, it is therefore interesting to look at the relative error $1 - f_{\mathrm{PT}}/f_{\mathrm{PA}}$ of the free energy per spin of parallel tempering against population annealing. The relative error is shown in Fig. 4.5. The errors are well bounded within the accuracy of about $10^{-4}$ and are well scattered around zero suggesting the nature of the errors is essentially statistical and again showing PT is as efficient as PA in accurately measuring the absolute free energy.

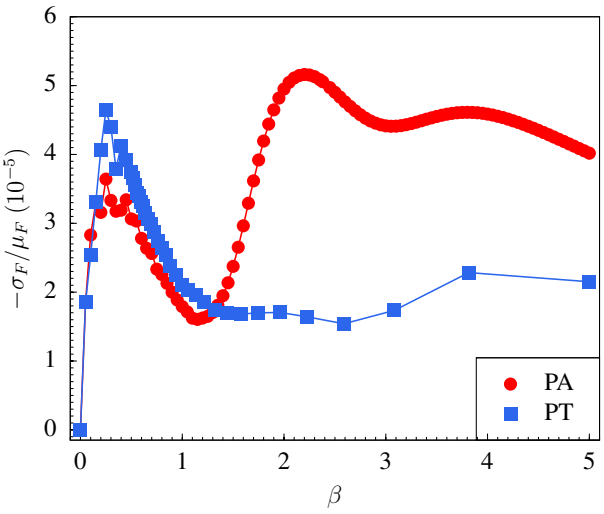


**Figure 4.2.** Evolution of the relative error of the estimated free energy $-\sigma_F/\mu_F$ as a function of the inverse temperature $\beta$ for a hard sample of $L = 8$. The data is the same as the largest runs of Fig. 4.1 for PA and PT. Note that the magnitude of the relative error does not grow as temperature is lowered.

## 4.4 Conclusions

In this chapter, I showed an efficient approach to measure the absolute free energy in spin-lattice models in parallel tempering using the free energy perturbation method. The algorithm is tested using the three-dimensional EA model and the results of parallel tempering agree very well with those of population annealing, showing parallel tempering can equally efficiently measure free energy as population annealing. There are technical issues with parallel tempering that nevertheless can be readily solved. The fact that the reweighting technique works so well suggests that many quantities including energy and free energy can be accurately estimated using the reweighting technique at many temperatures with little overhead in parallel tempering without using interpolation techniques. The work also poses an interesting question of a careful and systematic comparison of the efficiency of PA, PT as well as other techniques like the Wang-Landau algorithm in measuring free energy. Finally, I hope that this simple and direct access to free energy in parallel tempering can make spin glass and other relevant research fields potentially richer.

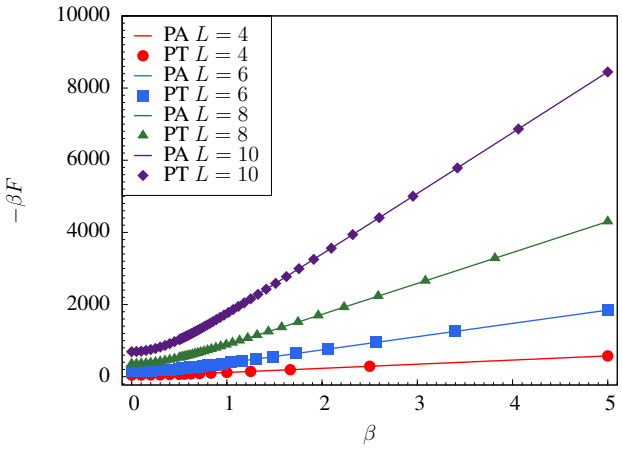


**Figure 4.3.** Comparison of $-\beta F$ for a typical sample of system sizes $L = 4, 6, 8$ and 10 in a wide range of temperatures. The PT data falls right on top of the PA curve, showing the effectiveness of PT in measuring free energy.

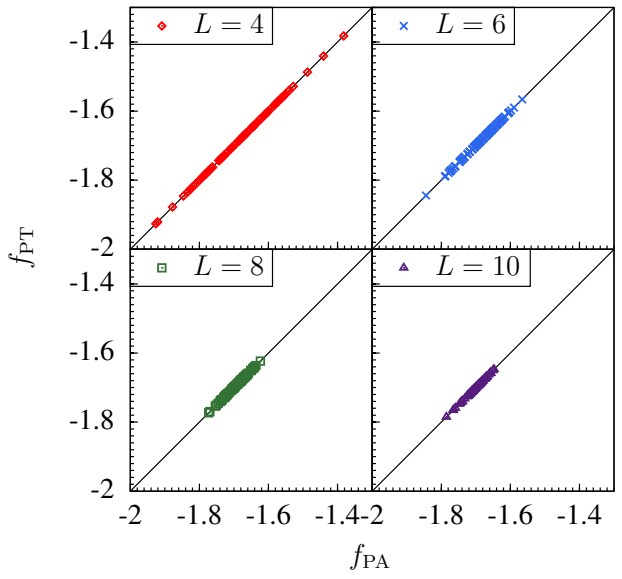


**Figure 4.4.** Scatter plot of free energy per spin $f$ of PA and PT of system sizes $L = 4, 6, 8$ and 10 at $T = 0.2$. Each point represents a sample.

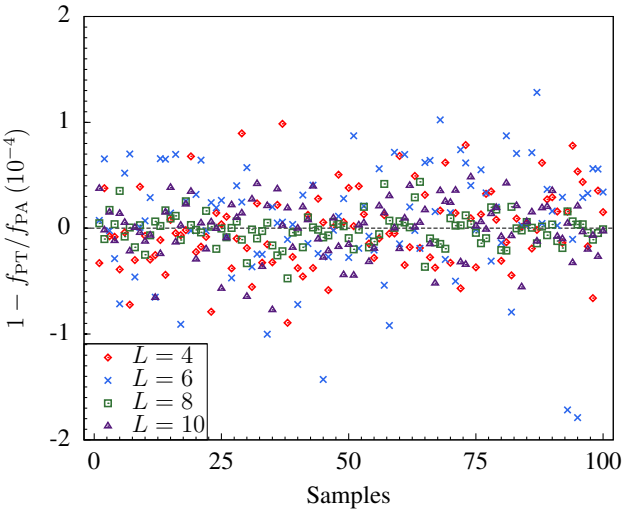


**Figure 4.5.** The same data as in Fig. 4.4, but plotting the relative error $1 - f_{\mathrm{PT}}/f_{\mathrm{PA}}$ of the free energy per spin of PT against PA at $T = 0.2$. Each point represents a sample.

# CHAPTER 5

# THERMAL BOUNDARY CONDITIONS

In this chapter, we will introduce thermal boundary conditions (TBC) and apply them to understanding the low temperature phase of the EA spin glass. In thermal boundary condition, in each direction at the boundaries, the system has freedom to choose periodic boundary conditions (PBC) or anti-periodic boundary conditions (APBC) according to Boltzmann weights. So in $3d$, there will be a total of 8 boundary conditions competing with one another. I will also introduce sample stiffness which measures the cost of inserting domain walls in a sample. The nature of the spin glass phase of the $3d$ EA model is then studied using TBC and sample stiffness extrapolation. This chapter is adapted from Ref. [125].

## 5.1 Thermal boundary conditions

The motivation for using thermal boundary conditions for spin glasses can be explained by considering two simpler examples–the ferromagnetic and antiferromagnetic Ising models on a square lattice with lattice size $L$ an odd number. For the ferromagnetic Ising model periodic boundary conditions in all directions are natural and appropriate because they do not induce domain walls in the ordered phase. However, for the antiferromagnetic Ising model, periodic boundary conditions will induce domain walls and the observables for finite systems will have strong finite size corrections. The natural boundary conditions for the antiferromagnet with $L$ odd are antiperiodic in all directions. Now, suppose we are asked to simulate an Ising model but we are not told whether it is a ferromagnet or antiferromagnet. If we use thermal boundary conditions then we will automatically choose the natural boundary conditions independent of which model we have been given,

namely periodic in all directions if the system is a ferromagnet and antiperiodic in all directions if the system is an antiferromagnet. The other boundary conditions will induce domain walls and therefore have higher free energies. The difference in free energy between thermal boundary conditions and any of the domain-wall-inducing boundary conditions scales as $L^\theta$ where $\theta$ is the spin stiffness exponent. For the Ising model (either ferromagnetic or antiferromagnetic) in the low temperature phase, $\theta = d - 1 \geq 0$ and, even for modest system sizes, thermal boundary conditions are essentially the same as the single natural boundary condition because all unfavorable choices are suppressed.

While one can a priori determine the optimal boundary conditions for simple systems such as ferromagnets and antiferromagnets, the same is not true for spin glasses. For a given sample, a single boundary condition such as PBC may induce domain walls and induce large finite size effects. The motivation for using thermal boundary conditions is thus the same as for the simple (anti)ferromagnetic example discussed above. Because we do not know which of the eight periodic/antiperiodic boundary conditions fits the sample best, we simply let the system choose by minimizing the free energy.

At zero temperature, thermal boundary conditions correspond to selecting from among the $2^d$ boundary conditions those with the lowest energy ground states. These boundary conditions have been employed with exact algorithms for finding ground states of two-dimensional spin glasses [67, 116]. TBC was called "extended" boundary conditions in Ref. [116] and was also argued to minimize finite size effects. Similar ideas but using periodic and antiperiodic boundary conditions in a single direction are discussed in [47, 107, 108, 44].

Thermal boundary conditions can also be used to measure the spin stiffness exponent $\theta$ by comparing the free energy of TBC with other boundary conditions. For example, for spin glasses it is sufficient to compare the free energy for TBC with that for PBC. This approach is expected to yield the same exponent but a different prefactor for the spin stiffness as compared to the standard

method of taking the absolute value of the free energy difference between periodic and antiperiodic boundary conditions.

## 5.2 Spin stiffness and sample stiffness

We measure the free energies, $F_{\mathcal{J}}^{\mathrm{TBC}}$ and $F_{\mathcal{J}}^{\mathrm{PBC}}$ for each sample $\mathcal{J}$ in thermal and periodic boundary conditions, respectively. We also measure the ground-state energies $E_{\mathcal{J}}^{\mathrm{TBC}}$ and $E_{\mathcal{J}}^{\mathrm{PBC}}$ for each sample in both TBC and PBC, respectively. We compute the ground-state energy by taking the minimum energy in the population at the lowest temperature ($T = 0.2 \ll T_c$) as discussed in Chapter 3.

The traditional measure of spin stiffness is the difference between the free energy, or at zero temperature, the ground-state energy of two different boundary conditions–usually periodic and antiperiodic in a single direction with periodic boundary conditions in all other directions. For spin glasses, this quantity may be of either sign and the absolute value must be taken before performing the disorder average. Here we consider the free energy (ground state energy) difference between thermal boundary conditions and periodic boundary conditions. This quantity is nonnegative because periodic boundary conditions are contained in the TBC ensemble of boundary conditions so no absolute value needs to be taken. We refer to $\Delta F$ as the disorder average free energy (ground-state energy) difference between TBC and PBC. The scaling of $\Delta F$ with system size $L$ defines the spin stiffness exponent $\theta$,

$$\Delta F \sim L^{\theta}. \tag{5.1}$$

We measure $\theta$ at $T = 0$, 0.2 and 0.42 by fitting to this equation.

The free energy for each boundary condition in the TBC ensemble can be measured by partitioning $Q$ into its eight boundary condition components but we did not collect data to do this measurement. Instead, we estimate the ratio of the free energy of the dominant boundary condi-

tion to the free energy of all the other boundary conditions combined. Let $f_{\mathcal{J}}$ be the fraction of the population in the boundary with the largest population in sample $\mathcal{J}$. The quantity $\lambda_{\mathcal{J}}$,

$$\lambda_{\mathcal{J}} = \log \frac{f_{\mathcal{J}}}{(1 - f_{\mathcal{J}})} \tag{5.2}$$

is an estimator of the free-energy difference (times $-\beta$) between the dominant boundary condition and all other boundary conditions in sample $\mathcal{J}$. Note that $\lambda_{\mathcal{J}}$ is a measure of the stiffness of sample $\mathcal{J}$. If only one boundary condition dominates the ensemble of boundary conditions it means that inserting a domain wall is very costly and the sample is stiff while if $\lambda_{\mathcal{J}}$ is small, the domain walls induced by changing boundary conditions have little cost and the sample is not stiff. Note that, in principle, all boundary conditions could have equal weight so $\lambda \geq -\log 7$. Samples with negative values of $\lambda_{\mathcal{J}}$ are not stiff.

## 5.3 EA model in TBC

In this section, we present results for the spin stiffness (5.3.1), the order parameter distribution near zero, $I_{\mathcal{J}}$, (5.3.2) and the correlation of $I_{\mathcal{J}}$ and $\lambda_{\mathcal{J}}$ (5.3.3). The main result of this section is that $\lambda_{\mathcal{J}}$ increases with system size and that stiff samples have small values of $I_{\mathcal{J}}$.

### 5.3.1 Spin stiffness

Figure 5.1 shows the free-energy difference or, for $T = 0$, ground-state energy difference $\Delta F$ between TBC and PBC for temperatures $T = 0$, 0.2 and 0.42 as a function of system size $L$. The straight lines are best fits to the functional form $\Delta F \sim aL^{\theta}$. The fits for $\theta$ are shown in Table 5.1. The result $\theta(T = 0) = 0.197 \pm 0.017$ is in reasonable agreement though at the low end of previous measurements of $\theta$ carried out at zero temperature [43, 90, 24, 13]. Note that the stiffness exponent has not previously been measured at nonzero temperature. We see that $\theta$ decreases as temperature increases. Presumably, this is a finite-size effect because $\theta$ is expected to have a single asymptotic value throughout the low temperature phase [37].

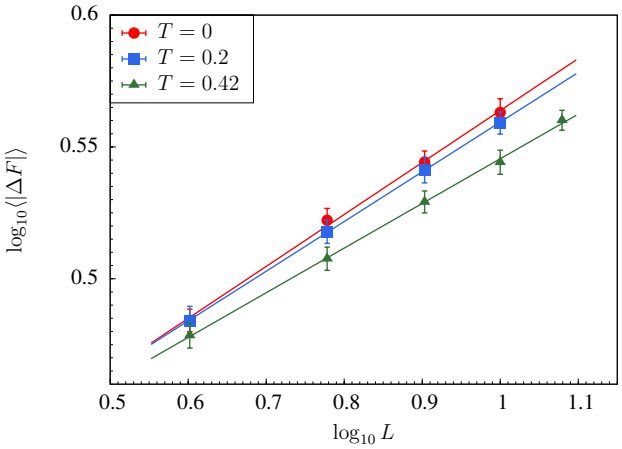


**Figure 5.1.** Free-energy change $\Delta F$ vs system size $L$ for $T = 0$, 0.2, and 0.42. The straight lines are fits of the form $\Delta F \sim aL^{\theta}$.

**Table 5.1.** Estimates of the stiffness exponents $\theta$ and $\theta_\lambda$ for different temperatures $T$.

| $T$ | 0 | 0.2 | 0.42 |
|---|---|---|---|
| $\theta$ | 0.197(17) | 0.189(17) | 0.169(12) |
| $\theta_\lambda$ | — | 0.290(30) | 0.268(20) |

Next consider the sample stiffness measure $\lambda_{\mathcal{J}}$, defined in Eq. (5.2). Let $G_L(\lambda)$ be the cumulative distribution function for $\lambda$. The left panel of Fig. 5.2 is a log plot of $1 - G_L(\lambda)$, the complementary cumulative distribution function, for $\lambda$ for $T = 0.42$, and sizes 4 through 12. The nearly straight line behavior of $\log(1-G_L(\lambda))$ is indicative of a nearly exponential tail and suggests a data collapse if $\lambda$ is scaled by a characteristic $\lambda_{\text{char}}(L)$ given by the slope of the line. Since the tail is not perfectly straight, we instead define $\lambda_{\text{char}}(L)$ in terms of median-like quantities. If the distribution were exactly exponential then $1-G(\lambda) = e^{-\lambda/\lambda_{\text{char}}}$ and $1-G(\lambda_{\text{char}} \log b) = 1/b$ for any $b$. If the distribution is not perfectly exponential, $\lambda_{\text{char}}$ depends on $b$ so there is some ambiguity in the definition. We choose $b$ such that $\lambda_{\text{char}}$ is obtained from the tail of the distribution but not so far into the tail that the statistics are poor. For $T = 0.2$ we choose $b = 2$ so $\lambda_{\text{char}}$ is defined as the median divided by $\log 2$. For $T = 0.42$ we choose $b = 10$ to ensure that $\lambda_{\text{char}}$ is obtained from the tail of the distribution. The right panels of Figs. 5.2 and 5.3 show $1 - G_L(\lambda/\lambda_{\text{char}}(L))$

for $T = 0.42$ and $T = 0.2$, respectively, and reveal that all of the cumulative distributions collapse onto the same curve when scaled by $\lambda_{\rm char}(L)$.

Figure 5.4 shows $\lambda_{\rm char}(L)$ vs $\log L$ for $T = 0.2$ and $T = 0.42$. Since $\lambda_{\rm char}(L)$ is a stiffness measure, we can extract a new stiffness exponent $\theta_\lambda$ from a fit to the form,

$$\lambda_{\rm char}(L) \sim aL^{\theta_\lambda}. \tag{5.3}$$

The values of $\theta_\lambda$, given in Table 5.1, are larger than $\theta$ obtained from the average free energy difference but close to the value, 0.27, found in Ref. [24] using aspect ratio scaling. Presumably, the asymptotic values of $\theta$ and $\theta_\lambda$ are the same. We prefer the larger value, $\theta_\lambda$ because it is obtained from the *tail* of the stiffness distribution so we believe it reflects the large-size behavior more accurately than the average free energy difference that defines $\theta$. Aspect ratio scaling is an independent way to minimize finite size effects and it is interesting that these two approaches yield the same answer within error bars.

It seems clear that $\lambda_{\rm char}(L) \to \infty$ as $L \to \infty$. At least on a coarse scale, the full distribution $G_L(\lambda)$ also scales with $\lambda_{\rm char}(L)$. A closer look at $G$ near the head of the distribution for $T = 0.42$ shows that the data collapse is not perfect and there are significant finite-size corrections near $\lambda = 0$. Figure 5.5 shows $1 - G_L(\lambda/\lambda_{\rm char}(L))$ vs $\lambda$ in the region near $\lambda = 0$ for $T = 0.42$. Note that $1 - G_L(\lambda/\lambda_{\rm char}(L))$ appears to be increasing with $L$. Figure 5.6 is the same plot for $T = 0.2$. Because $G_L(0)$ is so small for $T = 0.2$, the error bars are too large to discern whether there is a trend with $L$. A reasonable hypothesis is that there is an asymptotic $L \to \infty$ scaling function $G_\infty(z)$ where $z = \lambda/\lambda_{\rm char}$ such that $G_L(\lambda) \to G_\infty(\lambda/\lambda_{\rm char}(L))$. The straight line behavior of $\log(1 - G_\infty(z))$ for large $z$ and increasing trend with $L$ for small $z$ suggests that $G_\infty(0) = 0$ and $G_\infty(z)$ is exponential for $z \gg 1$. In more physical terms, if $G_\infty(z)$ exists and is zero for $z \to 0^+$, it means that a single boundary condition dominates the TBC ensemble almost surely, i.e., the dominant boundary condition almost always has a much lower free energy than the other seven

boundary conditions. A more complicated possibility is that $G_\infty(0^+) > 0$. The consequences of these possibilities for the RSB vs two-state pictures are discussed in Sec. 5.4.

It is noteworthy that for the system sizes accessible to simulations, $\lambda_{\rm char}(L)$ is sufficiently small that the TBC ensemble contains a mixture of several competing boundary conditions for a substantial fraction of samples. The disorder average $I_L$ is dominated by these samples and is therefore not characteristic of the large-$L$ behavior when $\lambda_{\rm char}(L)$ is expected to be large. In what follows we circumvent this difficulty by extrapolating first in $\lambda$ and then in $L$, making use of the fact our data contain a relatively large dynamic range in $\lambda$.

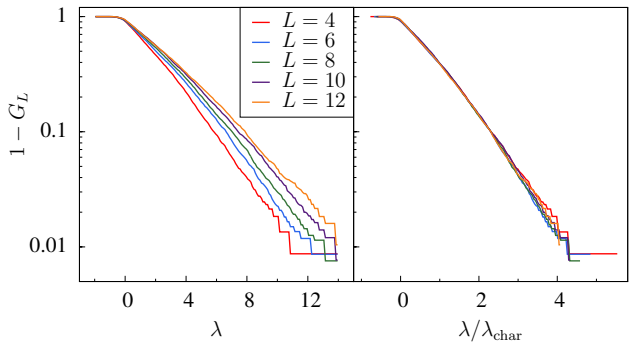


**Figure 5.2.** Left panel: Linear-log plot of $1 - G_L(\lambda)$ (the complementary cumulative distribution function) vs $\lambda$ for sizes $L = 4$ through 12 at $T = 0.42$. Right panel: $1 - G_L(\lambda/\lambda_{\rm char}(L))$ vs $\lambda/\lambda_{\rm char}(L)$.

### 5.3.2 Order parameter near $q = 0$

Figure 5.7 shows $I_L$, the disorder average of the integrated order parameter distribution with $|q| < 0.2$, as a function of size $L$ for temperatures $T = 0.2$ and 0.42, as well as for both PBC and TBC. For both boundary conditions, $I_L$ is, within error bars, independent of system size.

The constancy of $I_L$ has been taken as strong evidence for the RSB picture because the two-state picture predicts $I_L$ should decrease as $L^{-\theta}$. However, in what follows we argue that in TBC ultimately $I_L \to 0$ for very large $L$. On first glance the results for TBC are surprising since $I_L^{\rm TBC}$ is larger by more than a factor of two than $I_L^{\rm PBC}$. The explanation is that for many samples

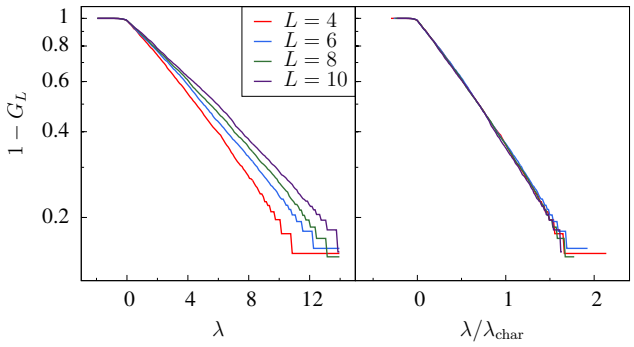


**Figure 5.3.** Left panel: Linear-log plot of $1 - G_L(\lambda)$ (the complementary cumulative distribution function) vs $\lambda$ for sizes $L = 4$ through 10 at $T = 0.2$. Right panel: $1 - G_L(\lambda/\lambda_{\rm char}(L))$ vs $\lambda/\lambda_{\rm char}(L)$.

the TBC ensemble contains several boundary conditions with significant weight and the overlap between spin configurations with different boundary conditions will tend to have small values of $q$ due to the existence of a relative domain wall. We shall return to this important point in Sec. 5.4.

### 5.3.3 Order parameter near $q = 0$ vs sample stiffness

Figure 5.8 is a scatter plot showing many of the disorder samples at $T = 0.42$ (left panel) and $T = 0.2$ (right panel) for all the sizes studied using TBC. Each point on the plot represents a sample $\mathcal{J}$. The $x$-coordinate of the point is $\lambda_{\mathcal{J}} = \log[f_{\mathcal{J}}/(1 - f_{\mathcal{J}})]$ and the $y$-coordinate is $I_{\mathcal{J}}$. Figure 5.9 is the same as Fig. 5.8 but with each system size on a separate plot for $T = 0.2$. The qualitative features of the plots are the same for each size although, as described above, the distribution of $\lambda$'s shifts to larger values for larger $L$. These figures together with the behavior of the $\lambda$-distribution constitute the main results of this chapter and motivate our conclusion that $I_{\mathcal{J}} \to 0$ almost surely as $L \to \infty$ in thermal boundary conditions.

Samples $\mathcal{J}$ for which $I_{\mathcal{J}}$ or $1 - f_{\mathcal{J}}$ are exactly zero within the precision of the simulations are not shown on these log-log plots since $\log I_{\mathcal{J}} = -\infty$ or $\lambda_{\mathcal{J}} = \infty$. The fractions of such samples are given in Table 5.2. Note that actual values of $I_{\mathcal{J}}$ or $1 - f_{\mathcal{J}}$ are are never exactly zero for finite systems; zeros correspond to values smaller than can be represented by the finite population sizes

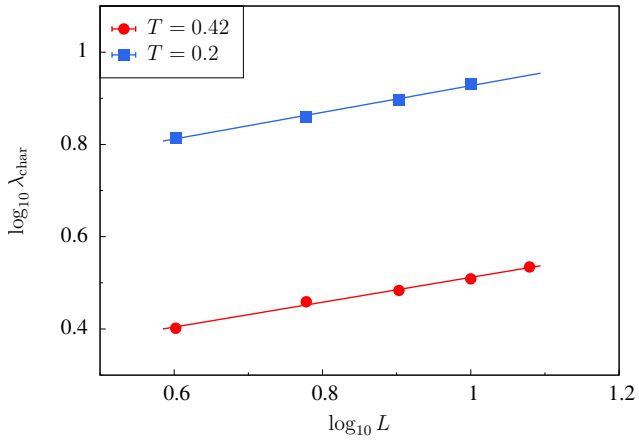


**Figure 5.4.** Log-log plot of $\lambda_{\rm char}(L)$ vs $L$ for $T = 0.2$ and $T = 0.42$. The straight lines represent fits of the form $\lambda_{\rm char}(L) \sim aL^{\theta_\lambda}$.

used in the simulations. It is important to note that the trends shown in Fig. 5.8 continue to hold for the large values of $\lambda$ that are omitted from this figure. Including all sizes, there are 216 samples with $1 - f_{\mathcal{J}} = 0$ for $T = 0.42$ and 2996 such samples for $T = 0.2$. Of these, only 7 samples for $T = 0.42$ and 38 for $T = 0.2$ are measured to have nonzero values of $I_{\mathcal{J}}$. The average value of $I$ for only those samples with $1 - f_{\mathcal{J}} = 0$ are $2.5 \times 10^{-8}$ and $10^{-4}$ for $T = 0.42$ and $T = 0.2$, respectively.

A striking feature of Figs. 5.8 and 5.9 is that there is a bounding curve that becomes a straight line for large $\lambda$ with most samples lying *below* that curve. Why are there are two classes of samples, with most samples below the curve and a few above it? We speculate that the samples below the curve have nonzero values of $I_{\mathcal{J}}$ as a result of the overlap between spin configurations with *different* boundary conditions. The contribution to the overlap between different boundary conditions in the TBC ensemble cannot exceed $2f(1 - f)$ in the limit $f \to 1$, so that if this is the primary mechanism producing small overlaps in sample $\mathcal{J}$ then $\log I_{\mathcal{J}} < (-\lambda_{\mathcal{J}} + \log 2)$. The straight lines in Fig. 5.8 are defined by $\log I = (-\lambda + \log 2)$. Thus, for most samples, we believe that the primary contribution to $I_{\mathcal{J}}$ comes from the overlap between different boundary conditions. For the rare samples above the bounding curve, the primary contribution to $I_{\mathcal{J}}$ must come from small overlaps within the dominant boundary condition.

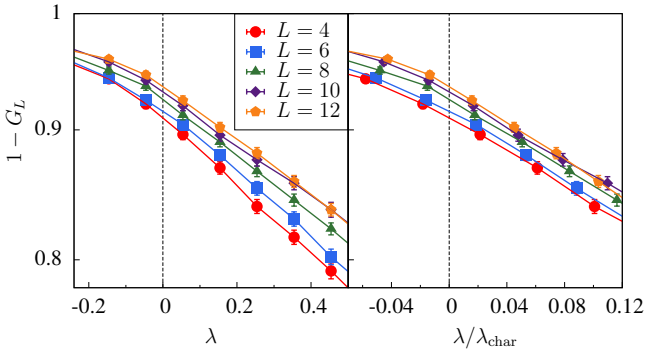


**Figure 5.5.** Left panel: $1 - G_L(\lambda)$ vs $\lambda$ for system sizes $L = 4$ through 12 at $T = 0.42$ in the region near $\lambda = 0$. Right panel: $1 - G_L(\lambda/\lambda_{\rm char}(L))$ vs $\lambda/\lambda_{\rm char}(L)$. Note that $1 - G_L(0)$ increases slowly with $L$.

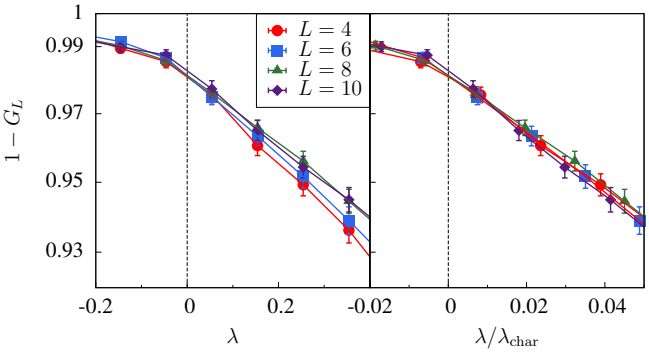


**Figure 5.6.** Left panel: $1 - G_L(\lambda)$ vs $\lambda$ for system sizes $L = 4$ through 10 at $T = 0.2$ in the region near $\lambda = 0$. Right panel: $1 - G_L(\lambda/\lambda_{\rm char}(L))$ vs $\lambda/\lambda_{\rm char}(L)$.

A second important feature of Figs. 5.8 and 5.9 is that the rare samples above the bounding curve have $I_{\mathcal{J}}$ roughly uniformly distributed on a logarithmic scale between the bounding curve and 1. On a linear scale this means that for large $\lambda$ almost all of these samples have small values of $I_{\mathcal{J}}$. Let $\rho(x|y)$ be the conditional probability density for $x = I_{\mathcal{J}}$ conditioned on $y = \lambda_{\mathcal{J}}$. If this distribution is exactly uniform above the bounding line then the part of the distribution above the line would take the form

$$\rho(x|y) = \frac{1 - w(y)}{xy} \qquad \text{for} \qquad x > 2\exp(-y), \tag{5.4}$$

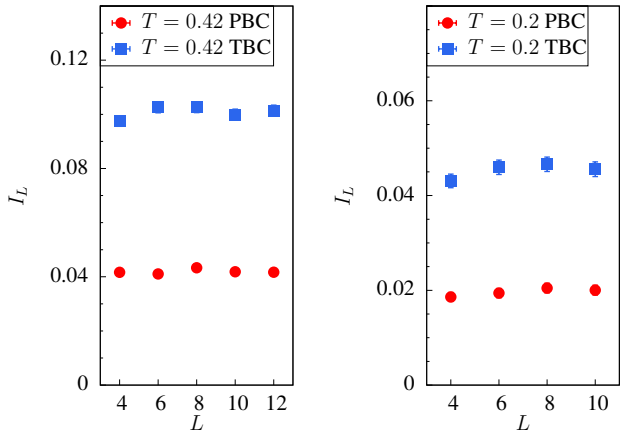


**Figure 5.7.** $I_L$ vs $L$ for PBC and TBC at temperature $T = 0.42$ (left panel) and $T = 0.2$ (right panel). The data seem independent of system size, suggesting an RSB interpretation of the data.

where $w(y)$ is the fraction samples below the bounding line. Figure 5.10 shows histograms of $I_{\mathcal{J}}$ values of the samples of all sizes that lie above the bounding line for the two temperatures. The position $\alpha$ along the $x$-axis is the scaled logarithmic distance between the bounding line and one. That is, $\alpha_{\mathcal{J}} = -\log I_{\mathcal{J}}/[\lambda - \log(2)]$ so that zero corresponds to large values, $I_{\mathcal{J}} \approx 1$ while one corresponds to $I_{\mathcal{J}}$ on the bounding line. For $T = 0.2$ the distribution is indeed relatively uniform on a logarithmic scale while for $T = 0.42$ it is skewed to small value of $I$.

## 5.4 Discussion

Three salient features of the data are apparent from Fig. 5.4, 5.8, and 5.10:

I. Typical values of the sample stiffness $\lambda_{\mathcal{J}}$ increase with system sizes $L$, as described by $\lambda_{\text{char}}(L)$.

II. Most samples have $I_{\mathcal{J}}$ less than a bounding curve described by $2e^{-\lambda}$ for large $\lambda$.

III. Samples with $I_{\mathcal{J}}$ above the bounding curve have $I_{\mathcal{J}}$ distributed more or less uniformly in $\log I$ between the bounding curve and one.

**Table 5.2.** Fraction of samples with $I_{\mathcal{J}} = 0$ and $f_{\mathcal{J}} = 1$ for different sizes, temperatures and boundary conditions.

| PBC | | | | | |
|---|---|---|---|---|---|
| System size $L$ | 4 | 6 | 8 | 10 | 12 |
| Fraction $I_{\mathcal{J}} = 0$ $(T = 0.42)$ | 0.21 | 0.19 | 0.16 | 0.16 | 0.19 |
| Fraction of $I_{\mathcal{J}} = 0$ $(T = 0.2)$ | 0.60 | 0.57 | 0.55 | 0.54 | – |
| Fraction of $f_{\mathcal{J}} = 1$ $(T = 0.42)$ | – | – | – | – | – |
| Fraction of $f_{\mathcal{J}} = 1$ $(T = 0.2)$ | – | – | – | – | – |

| TBC | | | | | |
|---|---|---|---|---|---|
| System size $L$ | 4 | 6 | 8 | 10 | 12 |
| Fraction $I_{\mathcal{J}} = 0$ $(T = 0.42)$ | 0.05 | 0.04 | 0.03 | 0.03 | 0.03 |
| Fraction of $I_{\mathcal{J}} = 0$ $(T = 0.2)$ | 0.35 | 0.33 | 0.30 | 0.28 | – |
| Fraction of $f_{\mathcal{J}} = 1$ $(T = 0.42)$ | 0.009 | 0.009 | 0.008 | 0.008 | 0.010 |
| Fraction of $f_{\mathcal{J}} = 1$ $(T = 0.2)$ | 0.15 | 0.16 | 0.15 | 0.15 | – |

We conjecture that these features hold for arbitrarily large $L$ and all temperatures in the low temperature phase. Assuming the above statements are asymptotically correct, we can draw some strong conclusions about how $I_L$ behaves for sufficiently large $L$ that $\lambda_{\text{char}}(L) \gg 1$. Note that these system sizes are far larger than are accessible in our simulations but, given the large dynamic range in $\lambda$ we can extrapolate to these sizes by first extrapolating in $\lambda$. For large $L$, $\lambda_{\mathcal{J}}$ is nearly always large according to Statement (I). Furthermore, due to Statements (II) and (III), $I_{\mathcal{J}}$ is, almost always small when $\lambda_{\mathcal{J}}$ is large. Thus $I_{\mathcal{J}}$ is nearly always small when $L$ is large. *This conclusion is the main result of our analysis. It is consistent with two-state pictures but not consistent with the RSB picture.*

In addition to being consistent with our data, these conjectures are quite plausible. Statement (I) asserts that $\lambda_{\mathcal{J}}$ is a measure of sample stiffness and that in the low temperature phase, almost all samples become stiff for large system sizes. As discussed above, Statement (II) asserts that in TBC large values of $I_{\mathcal{J}}$ arise mostly from the overlap between two different boundary conditions. Statement (III) asserts that the free-energy cost of a large excitation in the dominant boundary condition is more or less logarithmically distributed between $\lambda$ and 0.

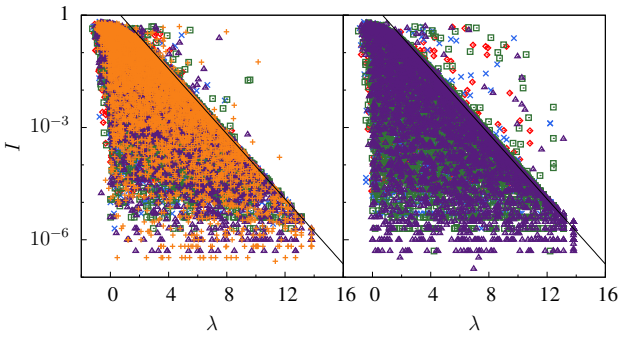


**Figure 5.8.** Scatter plots showing all disorder realizations for all system sizes at $T = 0.42$ (left panel) and $T = 0.2$ (right panel). Each point represents a sample $\mathcal{J}$ located at $x$-coordinate $\lambda_{\mathcal{J}}$ and $y$-coordinate $I_{\mathcal{J}}$. Red diamonds represent $L = 4$, blue crosses $L = 6$, green squares $L = 8$, purple triangles $L = 10$, and orange plus symbols $L = 12$.

We can make the arguments more quantitative using a simple model of how the disorder average $I_L$ will behave for large $L$. Let $\rho(x|y)$ be the conditional probability density for $I_{\mathcal{J}} = x$ conditioned on $\lambda_{\mathcal{J}} = y$. Based on Statements (II) and (III) we propose the form,

$$\rho(x|y) = w(y)\delta[x, 2\exp(-y)] + \frac{1 - w(y)}{xy}\theta[x - 2\exp(-y)], \tag{5.5}$$

where $w(y)$ is the fraction of samples at fixed $\lambda$ below the bounding curve, and $\theta(x)$ and $\delta(x, y)$ are the Heaviside function and the $\delta$-function, respectively. The first term conservatively places all the samples below the bounding curve on the curve itself. The second term represents the samples above the bounding curve with the conservative assumption that the distribution of $I_{\mathcal{J}}$ above the curve is uniform on a log scale as in Eq. (5.4). For purposes of the following calculation we assume that $w$ is a constant independent of $y$ but the qualitative conclusions do not depend on this assumption. Finally, we assume that the distribution of $\lambda$ obeys a size independent form $G_\infty(z)$ for the scaled variable $z = \lambda/\lambda_{\text{char}}(L)$. Note that we have assumed that the dependence of $I_L$ on $L$ is entirely through $\lambda_{\text{char}}(L)$ and that the conditional probability $\rho(x|y)$ is independent of $L$. These assumptions yield an explicit expression for $I_L$ as a function of $\lambda_{\text{char}}(L)$,

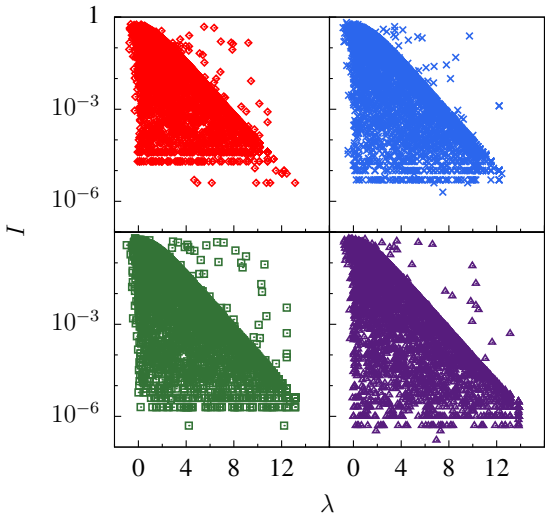


**Figure 5.9.** Same as Fig. 5.8 but for each system size in a separate panel and $T = 0.2$. Again, red diamonds represent $L = 4$, blue crosses $L = 6$, green squares $L = 8$, and purple triangles $L = 10$.

$$I_L = \frac{1}{\lambda_{\text{char}}(L)} \int_0^\infty dG_\infty(z) \int_0^\infty dx \; x\rho(x|z\lambda_{\text{char}}(L)). \tag{5.6}$$

Plugging in the ansatz of Eq. (5.5) and an exponential form for the scaled $\lambda$ distribution, $1 - G_\infty(z) = e^{-z}$, yields a somewhat complex expression involving exponential integrals whose asymptotic large $\lambda$ behavior is,

$$I_L \sim \frac{1}{\lambda_{\text{char}}(L)} \left[2w + (1-w)\log(\lambda_{\text{char}}(L))\right]. \tag{5.7}$$

Using Eq. (5.3) and assuming that asymptotically $\theta_\lambda = \theta$, we recover the prediction of the two-state picture that $I_L \sim L^{-\theta}$, however with a logarithmic correction that arises from the assumption of a log-uniform distribution for $I_{\mathcal{J}}$.

While the above assumptions lead to an explicit asymptotic expression for $I_L$ as a function $L$, this expression should not be taken too seriously. However the qualitative conclusion that $I_{\mathcal{J}} \to 0$ for almost all samples as $L \to \infty$ is robust and depends only on the asymptotic validity of Statements (I) – (III) above.

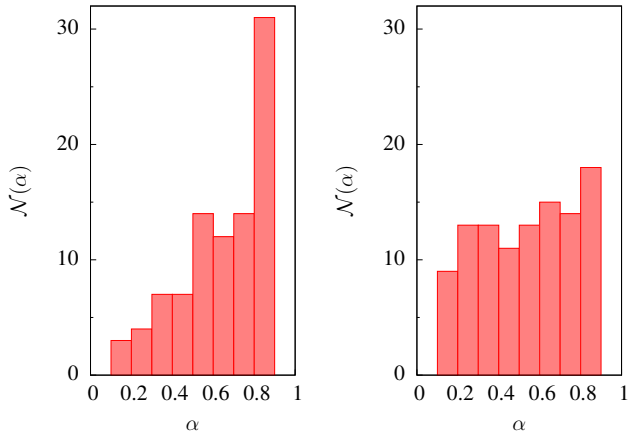


**Figure 5.10.** Histogram $\mathcal{N}(\alpha)$ for $\alpha_{\mathcal{J}} = \log(I_{\mathcal{J}})/[-\lambda_{\mathcal{J}} + \log(2)]$ for $T = 0.42$ (left panel) and $T = 0.2$ (right panel). $\alpha = 1$ corresponds to the small values of $I$ at the bounding line.

Given that $\lambda_{\text{char}}(L)$ is increasing with $L$ and that $I$ decreases with increasing $\lambda$, why is $I_L$ nearly constant for the sizes studied in our TBC simulations of the 3D EA model? We believe this conundrum can be explained, at least in part, by the fact that the main contribution to $I_L$ comes from samples with small values of $\lambda$. For $T = 0.2$, more than half the contribution to $I_L$ comes from samples with $\lambda < 1$ and more than 80% from $\lambda < 2$, and these fractions are even higher for $T = 0.42$. The head of the $\lambda$ distribution, has very little dependence on $L$, as can be seen in Figs. 5.2 and 5.3. Furthermore, the bounding curves in Fig. 5.8 are nearly flat in the small $\lambda$ region. Thus several effects come into play in keeping $I_L$ nearly independent of $L$. First, the main contribution to $I_L$ is from samples with small stiffness. Second, the fraction of samples with small stiffness decreases by only a small amount for the sizes studied and, finally, $I$ does not depend much on $\lambda$ for small $\lambda$. One would have to go to much larger sizes before $I_L$ would decrease according to the predicted asymptotic power law $L^{-\theta}$.

In the foregoing, we have assumed that $G_L(0) \to 0$ as $L \to \infty$ or, equivalently if $G_\infty(z)$ exists, $G_\infty(0^+) = 0$. We now consider the consequences of an alternate assumption that $\lambda_{\text{char}}(L) \to \infty$ and $G_\infty(z)$ exists but $G_\infty(0^+) > 0$. This possibility cannot be ruled out by the data although if it holds, it appears that $G_\infty(0^+)$ is quite small. In physical terms $G_\infty(0^+) > 0$ means that even

for very large sizes, a fraction $G_\infty(0^+)$ of samples has a mixed ensemble of boundary conditions in TBC while the remaining samples have only a single boundary condition in the TBC ensemble. This scenario would imply that the 3D EA model in TBC is divided into two classes of disorder realizations, one of which, with weight $(1 - G_\infty(0^+))$, has $I_L = 0$ and the other, with weight $G_\infty(0^+)$, has $I_L > 0$. This possibility seems unlikely but is not contradicted by the data. It has no straightforward explanation in either two-state or RSB pictures.

Our hypothesis is that thermal boundary conditions and periodic boundary conditions have the same behavior in the limit of large system sizes. We use thermal boundary conditions as a tool to improve the extrapolation to large system sizes from the very small system sizes accessible in simulations. It is known that coupling dependent boundary conditions are not equivalent to periodic boundary conditions and are not suitable for understanding properties of the spin glass phase because they could be used to select a single pure state even if coupling independent boundary conditions admit many pure states. The status of thermal boundary conditions with regard to coupling dependence is not clear. On the one hand, the TBC ensemble contains different mixtures of boundary conditions for different choices of couplings. On the other hand, the particular mixture of the eight boundary conditions is chosen by the system itself and is not imposed externally. As discussed in Sec. 5.1, our intuition is that TBC minimizes finite size effects rather introducing spurious physics but this question requires further investigation. In any case, we have provided compelling evidence that the $3d$ EA model in thermal boundary conditions is best described by a picture with a single pair of pure states in each finite volume.

## 5.5 Conclusion

We have introduced two new techniques with the aim of extrapolating to the large system-size behavior of finite-dimensional spin glasses at low temperature. First, we use thermal boundary conditions to minimize the effect of domain walls induced by boundary conditions. Second, we use a natural measure of sample stiffness defined within thermal boundary conditions and extrapolate

to large values of the sample stiffness. By noting that the sample stiffness increases with system size we then obtain an extrapolation in system size. The dynamic range in sample stiffness in the data is sufficiently large that a qualitative extrapolation is readily apparent. The conclusion is that nearly all large samples will have essentially no weight in the overlap distribution near zero overlap. The analysis also explains why this qualitative behavior cannot be seen using a direct extrapolation in system size for the small sizes studied. Our conclusions are consistent with two-state pictures but are inconsistent with the mean field, replica symmetry breaking picture.

Our results hold for thermal boundary conditions. We believe that thermal boundary conditions are equivalent to other coupling independent boundary conditions so that our conclusions about the infinite volume limit also apply to the more familiar periodic boundary conditions. However, it is important to investigate the equivalence of thermal and periodic boundary conditions. Furthermore, thermal boundary conditions and extrapolating in sample stiffness are general methods that should be useful in studying other finite-dimensional disordered systems.

# CHAPTER 6

# CHAOS IN SPIN GLASSES

In this chapter, we explore another application of thermal boundary conditions to spin glasses. We study the fragility of spin glasses to small temperature perturbations, known as temperature chaos. Temperature chaos is intrinsically related to the statistics of crossings in the free energy for different boundary conditions. By studying the energy difference between boundary conditions at free-energy crossings, we determine the domain-wall fractal dimension. Similarly, by studying the number of crossings, we determine the chaos exponent. Our results also show that computational hardness in spin glasses and temperature chaos are closely related. This chapter is adapted from Ref. [124].

## 6.1 Introduction to temperature chaos

Chaos refers to sensitivity to small perturbations. In addition to dynamical systems where the phenomenon was first identified, there are many statistical mechanical systems where chaotic effects have been predicted and observed. For example, the hysteresis, as well as memory and rejuvenation effects found in random elastic manifolds, polymers [38, 106, 25, 68], as well as spin glasses are considered to be a direct manifestation of the presence of chaos [89, 29, 52]. It is surprising and fascinating at the same time that the nonequilibrium and equilibrium states of spin glasses are so fragile to small perturbations. Chaos is therefore central to the understanding of both equilibrium and nonequilibrium properties of spin glasses, as well as related systems. Furthermore, there is mounting evidence that chaos in spin glasses is directly related to the computational hardness and long thermalization times [32] of these paradigmatic benchmark problems. As such, quantifying

and understanding chaotic effects in spin-glass-like Hamiltonians could be of great importance for the development of any novel algorithm or computing architecture [135, 55, 45].

In this chapter we study the effects of small thermal perturbations. *Temperature chaos* thus refers to the property that a small change in temperature results in a complete reorganization of the equilibrium configuration of the system. Temperature chaos has long been predicted for spin glasses [72, 97, 35, 20]. Although few early studies raised doubts about the existence of temperature chaos [10], increasing numerical evidence for temperature chaos has emerged in recent years for various models like the random-energy random-entropy model [65] and also more realistic three- and four-dimensional Ising spin glasses [107, 57, 32]. It has been suggested that temperature chaos would only be observable in spin glasses at very large system sizes and large changes in the temperature [3, 102]. However, some studies [57] demonstrated the existence of temperature chaos via scaling arguments. One direct manifestation of temperature chaos is that the free-energy difference between two boundary conditions that differ by a domain wall in the system may change sign as a function of temperature. Previous studies examined the free-energy difference between periodic and anti-periodic boundary conditions in a single direction to identify temperature chaos [115, 107]. This motivates us to study temperature chaos using *thermal boundary conditions* [125], in which all $2^d$ combinations of periodic and anti-periodic boundary conditions in the $d$ directions (space dimensions) appear in a single simulation with their appropriate statistical weights. Thermal boundary conditions provides an elegant way to study temperature chaos.

Here we quantitatively investigate this fascinating phenomenon using population annealing Monte Carlo. This simulation approach is ideal to study chaos effects in spin glasses because multiple boundary conditions can be studied at the same time. We show that temperature chaos is intrinsically related to the statistics of crossings in the free energy for a pair of boundary conditions [115] and thus establish both qualitatively and quantitatively the presence of chaos in spin glasses. Our approach can be applied to a multitude of problems and, in particular, to the search for hard benchmark instances for novel computing paradigms [55, 45].

What causes temperature chaos? Temperature chaos results from the existence of dissimilar classes of configurations with similar free energies but differing energies and entropies. Consider two classes of spin configurations, $\sigma_1$ and $\sigma_2$ corresponding to distinct basins in the free-energy landscape. Within each class, all spin configurations are similar but the two classes are dissimilar and differ by a large relative domain wall. Let $\Delta F(T)$ be the free-energy difference at temperature $T$ between these two classes, with $\Delta F(T) = \Delta E(T) - T\Delta S(T)$ where $\Delta E$ and $\Delta S$ are the energy and entropy, respectively, of the relative domain wall. Suppose now that $\Delta E$ and $\Delta S$ are both much larger than $\Delta F$ and weakly dependent on temperature; then a small change in temperature may lead to sign change in $\Delta F$. Suppose that $\Delta F$, $\Delta E$ and $\Delta S$ all behave as power laws in the size scale $\ell$ of the relative domain wall separating spin configurations $\sigma_1$ and $\sigma_2$ with leading behavior $\Delta F \sim \ell^{\theta}$ but with $\Delta E \sim \Delta S \sim \ell^{d_s/2}$ and $d_s/2 > \theta$. Here $\theta$ is the stiffness exponent and $d_s$ is the fractal dimension of the domain wall. As $\ell$ increases, the temperature perturbation $\delta T$ required to change the sign of $\Delta F$ decreases, i.e., $\delta T \sim \ell^{-\zeta}$ with the chaos exponent $\zeta$ given by $\zeta = d_s/2 - \theta$.

## 6.2 Probing temperature chaos using thermal boundary conditions

In the TBC ensemble each spin configuration and boundary condition pair occur in the ensemble with their respective Boltzmann weights. Thus, each boundary condition $i$ occurs in the ensemble with a weight depending on its free energy $F_i$. The probability $p_i$ of boundary condition $i$ in the ensemble is given by $p_i = \exp[-\beta(F_i - F)]$, where $F$ is the total free energy of the system in TBC and $\beta$ the inverse temperature.

In thermal boundary conditions, a domain wall on the scale of the linear system size $L$ separates each boundary condition. Thus temperature chaos manifests itself as a strong temperature dependence in the relative free energies of the different boundary conditions (BCs). Because the stiffness exponent is positive, in the low-temperature phase one expects that for large systems a single BC will dominate the ensemble for almost all temperatures. However, as the temperature changes, the dominant boundary condition will frequently change. A crossing event occurs when

the free-energy difference between two BCs changes sign. The proliferation of crossing events is a direct indication of temperature chaos. Boundary-condition crossing events between periodic and antiperiodic BCs in one direction were studied in the two-dimensional EA model in Ref. [115] and identified as a signature of temperature chaos. Figure 6.1 shows BC probabilities, $\{p_i\}$ for all eight boundary conditions as a function of temperature for a single $L = 10$ sample. As expected, at high temperatures, each BC occurs with equal probability. However, at low temperatures, four different BCs dominate in different temperature ranges and, indeed the dominant boundary condition at the lowest temperatures has a tiny probability in a range just below the critical temperature.

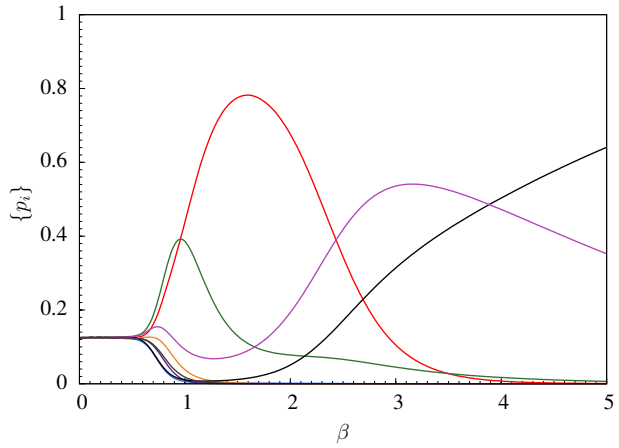


**Figure 6.1.** A single size $L = 10$ sample displaying several boundary-condition crossings. The plot shows the probability of the eight boundary conditions $\{p_i\}$ as a function of inverse temperature $\beta$.

We carried out simulation of the three-dimensional EA model in TBC using population annealing Monte Carlo. Thermal boundary conditions are easily simulated in population annealing by initializing the population at $\beta = 0$ with 1/8 of the population in each of the eight BCs. Resampling takes care of that at every temperature, each BC appears with the correct statistical weight. We study 2000 samples of sizes $L = 4$, 6, 8, 10 and 12 down to temperature $T = 0.33$, i.e., deep within the low-temperature phase.

The temperature difference between crossings scales as $L^{-\zeta}$ so that the number of crossing $N_C$ in a fixed temperature interval scales as $N_C \sim L^{\zeta}$. Also, at crossings, we have that $\Delta F = 0$ so that $d_s/2$ can be obtained from the scaling of the average of $\Delta E$ at crossings as a function of $L$. Finally, we measured the spin stiffness exponent in TBC in Chapter 5. Thus, within TBC we can independently measure all three exponents $\theta$, $d_s/2$ and $\zeta$ and verify the relation $\zeta = d_s/2 - \theta$.

Crossings can be divided into two classes: *Dominant* crossings are those such that the two equal BC probabilities at the crossing are larger than all other BC probabilities. All other crossings are *subdominant*. For large systems, the BC probability at a subdominant crossing is expected to be typically suppressed by $\exp(L^{-\theta})$ relative to the dominant BC and thus will be increasingly difficult to observe in TBC simulations. To avoid finite-size corrections in counting crossings, here we focus on dominant crossings. On the other hand, for measuring $\Delta E = T\Delta S$ ($\Delta S$ the change in entropy) we do not expect a distinction between dominant and subdominant crossings and, to improve statistics, we use all crossing with $p_i > 0.05$.

## 6.3 Results and discussions

Figure 6.2 shows a histogram of the number of crossings with $p_i > 0.05$ for four different system sizes as a function of inverse temperature. Figure 6.2 reveals that the number of crossings decreases with temperature, consistent with the fact that the entropy decreases with temperature so that increasingly large temperature changes are required to change the free-energy difference between BCs. Figure 6.2 also demonstrates that the density of crossings increases with system size (note that the bottom panels of Fig. 6.2 have a larger vertical scale). In the large-volume limit, the density of dominant crossings per sample is expected to become infinite.

Figure 6.3 is a log-log plot (base-10) of $N_C$ vs $L$ counting all dominant crossing in the range $\beta \in (1.5, 3.0)$. A simple power-law fit $N_C \sim L^{\zeta}$ yields $\zeta = 0.96(5)$. To test the effect of temperature on this exponent, we also calculated $\zeta$ from two smaller temperature ranges. For $\beta \in (1.5, 2.0)$ we find $\zeta = 1.05(8)$, and from $\beta \in (2.0, 3.0)$ we find $\zeta = 0.85(8)$. For higher temperatures,

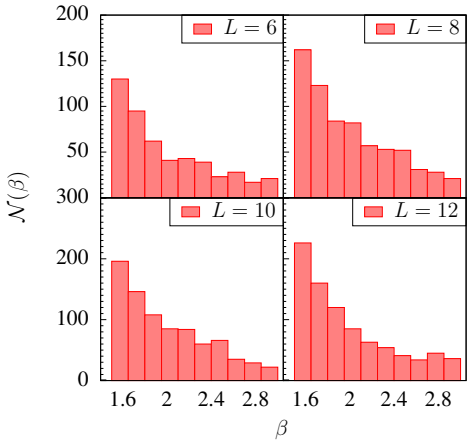


**Figure 6.2.** Distribution of the number of crossings with $p_i > 0.05$ with respect to $\beta$ for system sizes of $L = 6$, 8, 10 and 12.

critical fluctuations might contaminate the measurement of the chaos exponent while for lower temperatures the number of crossings is suppressed by the smallness of the entropy. We note that there is a significant trend to a smaller value of $\zeta$ at lower temperatures. If one assumes that a single exponent holds throughout the low-temperature phase, this trend suggests significant temperature-dependent finite-size corrections.

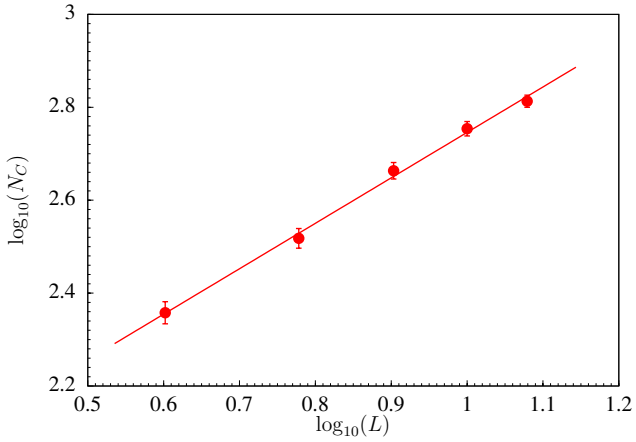


**Figure 6.3.** Number of dominant crossing in the range $\beta \in (1.5, 3.0)$ vs size $L$, for $L = 4$, 6, 8, 10, and 12. The straight line is the best power law fit (see text).

Figure 6.4 is a log-log plot (base-10) of the median and mean of the absolute energy difference $|\Delta E|$ vs $L$ at all crossings in the range $\beta \in (1.5, 3.0)$ such that $p_i > 0.05$. A simple power-law fit for the mean yields $|\Delta E| \sim L^{d_s/2}$ with $d_s/2 = 1.18(2)$. We again test the effect of the temperature range on $d_s/2$ by dividing the $\beta$ range into two intervals, $\beta \in (1.5, 2.0)$ and $\beta \in (2.0, 3.0)$ from which we obtain the the results $d_s/2 = 1.14(2)$ and $d_s/2 = 1.26(3)$, respectively. There is a significant trend toward larger values at lower temperatures, suggesting temperature-dependent finite-size corrections.

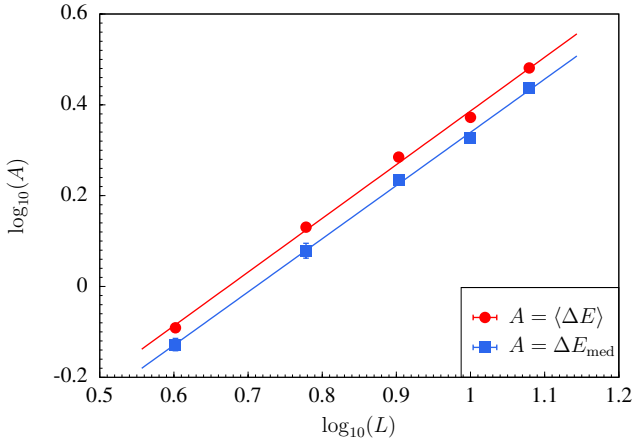


**Figure 6.4.** Mean and median energy difference between boundary condition crossings with $p_i > 0.05$ in the range $\beta \in (1.5, 3.0)$ for $L = 4$, 6, 8, 10, and 12. The straight line is the best power law fit.

Our results for the three-dimensional EA model, $d_s/2 = 1.19(4)$ and $\zeta = 0.95(5)$, are comparable but slightly smaller than previous work: For example, $d_s/2 = 1.29(1)$ was found in Ref. [91] based on perturbations of the ground state, and $d_s/2 = 1.31(1)$ was found in Ref. [58] based on the variance of the link overlap, while $\zeta = 1.04$ was found in Ref. [57] from the spin overlap between different temperatures. Combined with the estimate of $\theta = 0.27$ we find that the predicted relation $\zeta = d_s/2 - \theta$ is reasonably-well satisfied by our results from the full temperature range.

Temperature chaos partially explains why spin glass simulations are computationally costly [32]. All known efficient algorithms for equilibrating three-dimensional spin glasses rely on coupled sim-

ulations at many temperatures. Algorithms in this class include parallel tempering Monte Carlo [49], population annealing [48, 70], and the Wang-Landau algorithm [121]. In these algorithms, fast mixing at high temperatures provides new configurations to the low-temperature simulations. Temperature chaos decreases the effectiveness of these algorithms because the configurations supplied from higher temperatures are often rather different from the important configurations at lower temperatures. In TBC, this phenomenon means that BCs that are important at high temperature are unimportant at low temperature. This phenomenon is evident in Fig. 6.1. One might worry that boundary conditions that should be important at low temperature are completely lost at higher temperatures so that the simulations do not reach the correct TBC equilibrium. To verify that this is not the case, we performed an additional check of the equilibration of the TBC ensemble by re-doing several hundred of the hardest $L = 12$ samples using an order of magnitude larger population sizes in the simulation and we found no difference in the number of crossings for any sample.

A direct measure of hardness for a given sample is the entropic family size $\rho_s$, defined in Chapter 2. Figure 6.5 shows the disorder average of $\log \rho_s$ vs $L$ for two different classes of disorder samples. The $N_C = 0$ class has no temperature chaos events (crossings) in the range $\beta = 1.5$ to $\beta = 3$ while the $N_C > 0$ class has one or more temperature chaos events in the same range. The error bars are smaller than the data points and the curves show that $\rho$ scales exponentially in $L$ but that the exponential growth rate is faster for those samples with temperature chaos. It is an interesting question whether temperature chaos slows down all algorithms for spin glasses, not just those that depend on coupling multiple temperatures. However, studies have shown that [55, 45] computationally hard instances for classical algorithms are also computationally hard for quantum annealing machines, like the D-Wave Two quantum annealer. As such, by measuring $\rho_s$ for a given sample, we have a simple way to uniquely classify the complexity of a given instance. This means that our approach is of great importance in the development of hard problems to discern if quantum annealing can outperform simulated annealing simulations [112, 16, 15, 99, 104].

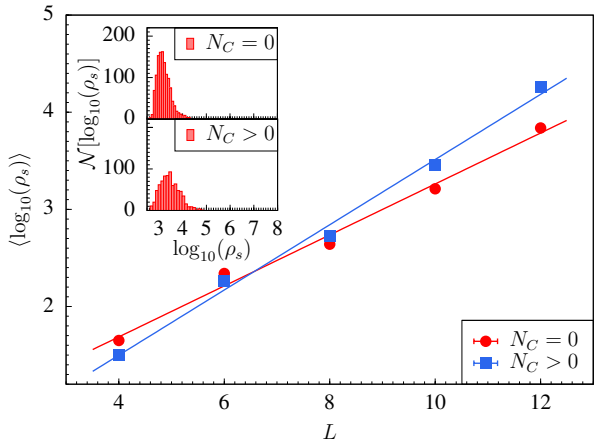


**Figure 6.5.** The average of the log of the hardness $\rho_s$ vs size $L$ for two classes of samples, those without crossing, $N_C = 0$ and those with at least one crossing, $N_C > 0$.

## 6.4 Bond chaos

The technique of thermal boundary conditions can also be adapted to the study of bond chaos. In this setting, one can first simulate a bond configuration $\{J^0_{ij}\}$ down to a low temperature like $\beta = 2$. Then one need to generalize the population annealing algorithm to do resampling with respect to changes in the Hamiltonian itself instead of the temperature. But this is straightforward when viewing population annealing as an algorithm that transform distributions.

### 6.4.1 Generalize the PA algorithm

In the regular PA algorithm, we have a set of equilibrium states at $\beta$ and would like to have a set of equilibrium states at $\beta'$. If replica $i$ has energy $E_i$, then the expected number of copies of replica $i$ is $\rho_i = \exp[-(\beta' - \beta)E_i]/Q$, where the normalization $Q = \dfrac{1}{R_0}\sum_i \exp[-(\beta' - \beta)E_i]$, where $R_0$ is the expected population size. Now suppose we have a set of equilibrium states at $\beta$ for bond configuration $J$ and we would like to transform the ensemble to a set of equilibrium states of a nearby bond configuration $J'$. A similar process can be carried out if we let $\rho_i = \exp[-\beta(E'_i - E_i)]/Q$, where the normalization $Q = \dfrac{1}{R_0}\sum_i \exp[-\beta(E'_i - E_i)]$. Here $E'_i$ and $E_i$ are the energy of replica $i$ for the bond configurations of $J'$ and $J$ respectively.

One can summarize the two different situations together. For two different sets of nearby distributions with weights $e^{-\mathcal{H}_i}$ and $e^{-\mathcal{H}'_i}$ for state $i$ respectively, then $\rho_i = \exp[-(\mathcal{H}'_i - \mathcal{H}_i)]/Q$ when changing the distribution from $\mathcal{H}$ to $\mathcal{H}'$, where the normalization $Q = \frac{1}{R_0}\sum_i \exp[-(\mathcal{H}'_i - \mathcal{H}_i)]$.

### 6.4.2 Scaling properties of bond chaos

Suppose the free energy cost to flip a droplet of size $\ell$ scales as $\ell^\theta$ at temperature $\beta$ and $J$, and the free energy cost to perturb the bonds with $\delta J$ for the droplet and the flipped droplet is $\Delta F_1$ and $\Delta F_2$ respectively. Then the free energy cost to flip the droplet at $J'$ is $\ell^\theta + \Delta F_2 - \Delta F_1$. One can easily see the effect of bond change for the last two terms is non-zero only at the surface of the droplet due to spin flip symmetries. Since $\Delta F = \Delta E - T\Delta S$ and the entropy change is much smaller than the energy change in the low temperature spin glass phase for bond chaos, one can conclude that $\Delta F_2 - \Delta F_1 \sim \ell^{d_s/2}$. Therefore, the energy cost to flip the droplet at $J'$ scales as $\Delta F' \sim \ell^\theta - \delta J \ell^{d_s/2}$ when there is a droplet flip. Therefore, the strength of bond change of $\delta J$ required for flipping a droplet of size $\ell$ scales as

$$\delta J \sim \ell^{-\zeta}, \tag{6.1}$$

where $\zeta = d_s/2 - \theta$.

### 6.4.3 TBC for bond chaos

Following the idea of the last two sections, we can generalize the techniques of temperature chaos to bond chaos. We choose an independent perturbation bond configuration $J'$ and and continuously transform $J_0$ to $J$ with a parameter $c$ as

$$J = \frac{J^0 + cJ'}{\sqrt{1+c^2}}. \tag{6.2}$$

Note that $J$ remains Gaussian distributed with mean 0 and variance 1 for all $c$. We can study bond chaos using boundary condition crossings by varying $c$. The scaling exponents of $\theta, d_s/2$ and

$\zeta$ can be therefore measured in a similar way. The similarities and differences of temperature chaos and bond chaos can also be investigated. A plot of the evolution of the weights $\{p_i\}$ for a typical sample of size $L = 8$ is shown in Fig. 6.6. The work is currently in progress.

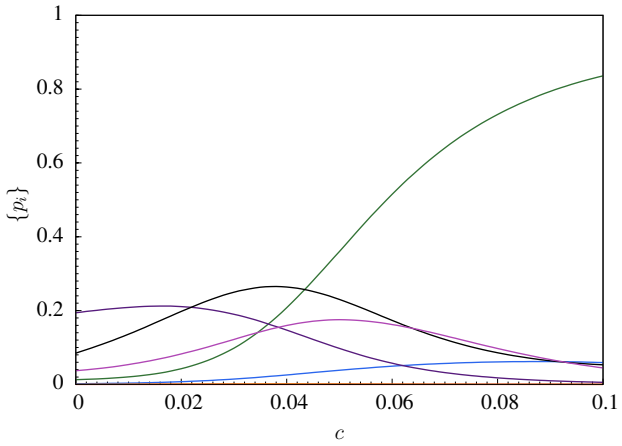


**Figure 6.6.** A typical $L = 8$ sample displaying several boundary-condition crossings. The plot shows the probability of the eight boundary conditions $\{p_i\}$ as a function of the tune parameter $c$.

## 6.5 Simulating TBC using parallel tempering

Finally, in this section, I briefly discuss two efficient methods to simulate thermal boundary conditions using parallel tempering, the diffusion method and the weighted average method.

### 6.5.1 The diffusion method

The most straight forward way to simulate thermal boundary conditions using parallel tempering is to include a set of high temperatures, including the infinite temperature $\beta = 0$, where each state and boundary condition appears equally likely. One can then simply propose random states and boundary conditions at $\beta = 0$ and let them diffuse downwards using the swap moves of parallel tempering. The convenience of working at $\beta = 0$ is because the acceptance fraction of boundary condition changes at finite temperatures will be low since many bonds are affected when boundary conditions change. Therefore, the temperature $\beta = 0$ is a natural choice.

The implementation of this method is simple and detailed balance is satisfied, so the question is whether the method is efficient. This method is also very similar in spirit with the usual "competing annealing" approach of population annealing. I will show in Sec. 6.5.3 that the performance of the two methods are indeed similar, which justifies the simple but yet efficient diffusion method.

### 6.5.2 The weighted average method

The simple implementations of both parallel tempering and population annealing will eventually suffer problems for hard samples with strong temperature chaos. This effect is reflected as boundary condition crossings when temperature changes. This mechanism will lead to a die off of a large number of independent families in population annealing and create bottlenecks in temperature diffusion in parallel tempering. Systematic errors as well as statistical errors are therefore increased for both algorithms, and thus requiring a large population size in population annealing or a long running time in parallel tempering to achieve thermal equilibrium. As a result, it is useful to study a weighted average method which simulates all the 8 boundary conditions independently in parallel and the results can be combined using weighted average at the end of the simulation.

The free energy is needed to do the weighted average and the free energy can be measured very accurately using the free energy perturbation method in parallel tempering as discussed in Chapter. 4. It is easy to see that the average energy and free energy are respectively given by

$$E = \sum_i E_i p_i, \tag{6.3}$$

$$F = \sum_i F_i p_i + T \sum_i p_i \log p_i, \tag{6.4}$$

where $p_i = \dfrac{e^{-\beta F_i}}{\sum\limits_j e^{-\beta F_j}}$. The last term in the free energy average is from the macroscopic contribution of entropy. The overlap distribution can also be averaged as

$$P(q) = \sum_{ij} P_{ij}(q) p_i p_j, \tag{6.5}$$

where $P_{ij}(q)$ is the overlap distribution between boundary conditions $i$ and $j$.

The weighted average method itself is interesting in its own right as it provides a lot of information about a sample. For simulating thermal boundary conditions, the method can be used for samples that are very chaotic and hard to equilibrate using the two simple methods. When very strong temperature chaos occur, this method can be significantly more efficient and accurate as the time scale to diffuse through multiple bottlenecks in the diffusion method can be very long. The weighted average method is also well suited for parallel computing. In my implementation of this method, I have used OpenMP parallel computing with 8 cores and each core separately working on a boundary condition. Note that the idea of the weighted average also applies to population annealing, but it can be very costly in memory to keep replicas of each boundary condition separately. It is an advantage of parallel tempering over population annealing in that such simulations is not as expensive in memory.

Finally, it is worth noting that the weighted average method in general is not really 8 times computationally costly as the diffusion method for at least two reasons. First of all, simulating thermal boundary conditions is usually *more expensive* than simulating a single boundary condition. From the studies of population annealing and the similar performance of population annealing and parallel tempering, this is usually about $2 \sim 3$ times harder. Secondly, the weighted average does not require as many Monte Carlo sweeps and requires no swap moves at high temperatures where the free energy landscape is not rough, and indeed the only purpose of the high temperature Monte Carlo sweeps is to get access to the absolute free energy to do weighted average. As a matter of fact, the number of sweeps one spends at high temperatures do not grow with system size and the amount of work at high temperatures compared with low temperatures vanishes in the thermodynamic limit. On the other hand, the Monte Carlo sweeps at high temperatures are necessary for the efficient diffusion of boundary conditions in the diffusion method.

### 6.5.3 Results

I will now briefly compare the overlap distribution and the evolution of $\{p_i\}$ as a function of $\beta$ for the hardest sample of $L = 8$ of thermal boundary conditions in Chapter. 5. The results of two runs using the diffusion method and population annealing with about the same amount of work are shown in Figs. 6.7 and 6.8 respectively. Similar results for the weighted average method and population annealing are shown in Figs. 6.9 and 6.10. Results of different methods agree very well within statistical errors. Additional details of the three methods can be found in Ref. [122].

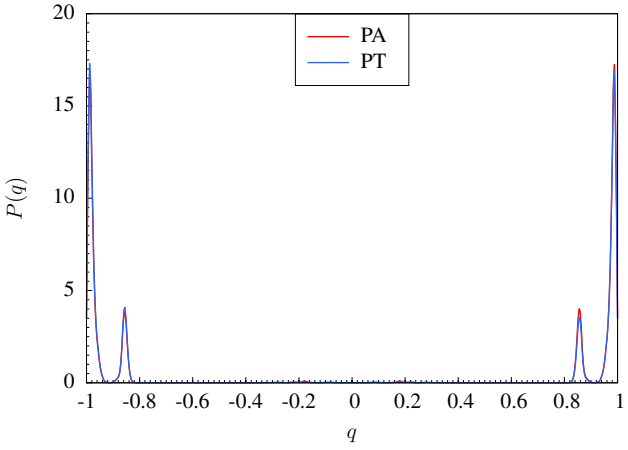


**Figure 6.7.** The comparison of the overlap distribution of the diffusion method and population annealing at $\beta = 5$.

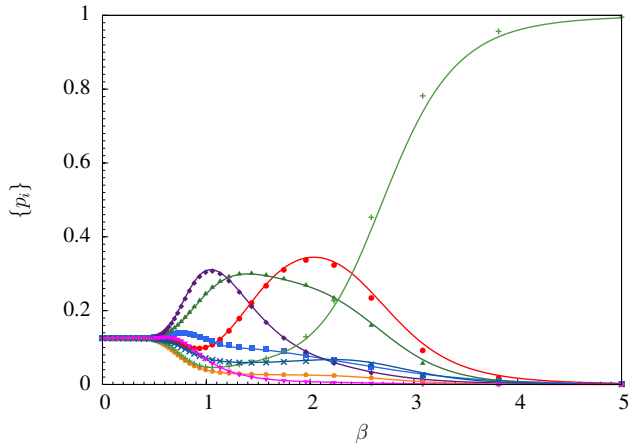


**Figure 6.8.** The comparison of the evolution of $\{p_i\}$ as a function of $\beta$ between the diffusion method and population annealing. The two methods predicts essentially the same set of $\{p_i\}$ at all temperatures.

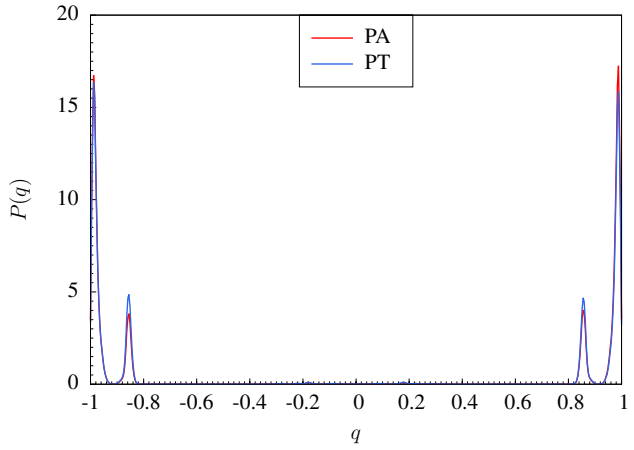


**Figure 6.9.** The comparison of the overlap distribution of the weighted average method and population annealing at $\beta = 5$.

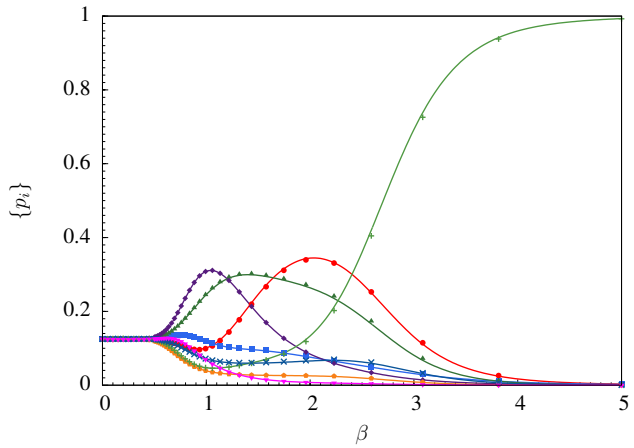


**Figure 6.10.** The comparison of the evolution of $\{p_i\}$ as a function of $\beta$ between the weighted average method and population annealing. The two methods predicts eseentially the same set of $\{p_i\}$ at all temperatures.

# CHAPTER 7

# SUMMARY AND FUTURE CHALLENGES

In this thesis, I have discussed the characterization and performance of an efficient algorithm, population annealing, and compared it with parallel tempering and simulated annealing. We found population annealing is similar in performance compared with parallel tempering in many aspects including thermal equilibrium sampling of states and optimization problems. Each algorithm has its own strengths and weaknesses. Population annealing has the advantage of being massively parallel and one can do weighted averaging to combine different independent runs to improve both statistical and systematic errors. Parallel tempering does not require much memory and converges exponentially while population annealing converges as a power law. Both algorithms outperform simulated annealing for finding spin glass ground states.

Secondly, we have studied in detail a boundary condition called thermal boundary conditions which reduces domain-wall effects. In combination with sample stiffness extrapolation, we show that our data is in agreement with a two-state scenario, not the RSB scenario. We have also applied TBC to the study of temperature chaos and bond chaos by studying scaling properties at boundary condition crossings and statistics of boundary condition crossings.

Our work has opened a number of new directions that are worth exploring. The first is the optimization of the population annealing algorithm in terms of tuning the annealing schedule, changing the population size as a function of temperature, using the cluster moves at intermediate temperatures and the kinetic Monte Carlo method at low temperatures. The comparison of the convergence of systematic errors of a single large run vs many small runs using the weighted average with the same amount of work is interesting for further studies. It is also interesting to test with

greater precision of the convergence properties of population annealing in exactly solvable systems like a one-dimensional spin-lattice model where the system can be numerically exactly solved using the transfer matrix technique. Applications of the population annealing algorithm to other hard statistical mechanical problems like the hard sphere gas, and the use of thermal boundary conditions for other disordered statistical mechanical systems like the random field Ising model should also be investigated. The MPI implementation for large scale optimization problems using population annealing is also going to be potentially useful.

It is interesting to check whether the same correlation features of sample stiffness $\lambda$ and the weight of the overlap distribution near zero $I$ persists in other short-range spin glass phases, in particular, the $4d$ EA model. A similar study of temperature chaos and bond chaos of the $4d$ EA model is interesting too. Finally, from the overlap of TBC, one can naturally ask how the disorder average of the weighted $I$ or the minimum $I$ of all the 8 boundary conditions changes with system size. In this way the spin overlap between different boundary conditions will not present. Some of this work is currently in progress.

# BIBLIOGRAPHY


[1] University of Cologne spin glass server, http://www.informatik.uni-koeln.de/spinglass/.

[2] R. Alvarez Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, M. Guidetti, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccione, and D. Yllanes. Nature of the spin-glass phase at experimental length scales. *J. Stat. Mech.*, P06026, 2010.

[3] T. Aspelmeier, A. J. Bray, and M. A. Moore. Why Temperature Chaos in Spin Glasses Is Hard to Observe. *Phys. Rev. Lett.*, 89:197202, 2002.

[4] F Barahona. On the computational complexity of Ising spin glass models. *J. Phys. A: Math. Gen.*, 15:3241, 1982.

[5] Andrei Băutu and Elena Băutu. Searching ground states of Ising spin glasses with genetic algorithms and binary particle swarm optimization. In N. Krasnogor, G. Nicosia, M. Pavone, and D. Pelta, editors, *Nature Inspired Cooperative Strategies for Optimization (NICSO 2007)*, volume 129 of *Studies in Computational Intelligence*, pages 85–94. Springer Berlin Heidelberg, 2008.

[6] Charles H Bennett. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.

[7] Bernd A. Berg and Thomas Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.*, 68:9–12, 1992.

[8] A. Billoire, L. A. Fernandez, A. Maiorano, E. Marinari, V. Martin-Mayor, G. Parisi, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, and D. Yllanes. Comment on "Evidence of Non-Mean-Field-Like Low-Temperature Behavior in the Edwards-Anderson Spin-Glass Model". *Phys. Rev. Lett.*, 110:219701, 2012.

[9] A. Billoire, A. Maiorano, E. Marinari, V. Martin-Mayor, and D. Yllanes. The cumulative overlap distribution function in realistic spin glasses. (arXiv:cond-mat/1406.1639), 2014.

[10] A. Billoire and E. Marinari. Evidence against temperature chaos in mean-field and realistic spin glasses. *J. Phys. A*, 33:L265, 2000.

[11] K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts and open questions. *Rev. Mod. Phys.*, 58:801, 1986.

[12] E. Bittner, A. Nußbaumer, and W. Janke. Make life simple: Unleash the full power of the parallel tempering algorithm. *Phys. Rev. Lett.*, 101:130603, 2008.

[13] S. Boettcher. Stiffness exponents for lattice spin glasses in dimensions $d = 3, \ldots, 6$. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(1):83–91, 2004.

[14] S. Boettcher and A. G. Percus. Optimization with extremal dynamics. *Phys. Rev. Lett.*, 86:5211–5214, 2001.

[15] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar. Experimental signature of programmable quantum annealing. *Nat. Comm.*, 4:2067, 2013.

[16] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer. Evidence for quantum annealing with more than one hundred qubits. *Nat. Phys,*, 10:218, 2014.

[17] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado. Strong Resilience of Topological Codes to Depolarization. *Phys. Rev. X*, 2:021004, 2012.

[18] H. Bombin and M. A. Martin-Delgado. Topological Quantum Distillation. *Phys. Rev. Lett.*, 97:180501, 2006.

[19] A. J. Bray and M. A. Moore. Scaling theory of the ordered phase of spin glasses. In L. Van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics and Optimization*, page 121. Springer, New York, 1986.

[20] A. J. Bray and M. A. Moore. Chaotic Nature of the Spin-Glass Phase. *Phys. Rev. Lett.*, 58:57, 1987.

[21] A. J. Bray and M. A. Moore. Scaling theory of the ordered phase of spin glasses. In L. Van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics*, volume 275 of *Lecture Notes in Physics*, page 57. Springer Verlag, 1987.

[22] V. Cannella, J. A. Mydosh, and J. I. Budnick. Magnetic susceptibility of Au-Fe alloys. *J.Appl. Phys*, 42:1689–1990, 1971.

[23] John L. Cardy, editor. *CURRENT PHYSICS SOURCES AND COMMENTS Vol.2 Finite-Size Scaling*. Elsevier Science Publishers B.V., 1988.

[24] A.C. Carter, A.J.Bray, and M.A. Moore. Aspect-Ratio Scaling and the Stiffness Exponent $\theta$ for Ising Sping Glasses. *Phys. Rev. Lett.*, 88:077201–1–077201–4, 2002.

[25] R. A. da Silveira and J.-P. Bouchaud. Temperature and Disorder Chaos in Low Dimensional Directed Paths. *Phys. Rev. Lett.*, 93:015901, 2004.

[26] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill. Topological quantum memory. *J. Math. Phys.*, 43:4452, 2002.

[27] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, F. Cioata, T. Cirip, P. Debuen, M. Drew-Brook, C. Enderud, S. Gildert, F. Hamze, J. P. Hilton, E. Hoskinson, K. Karimi, E. Ladizinsky, N. Ladizin-

sky, T. Lanting, T. Mahon, R. Neufeld, T. Oh, I. Perminov, C. Petroff, A. Przybysz, C. Rich, P. Spear, A. Tcaciuc, M. C. Thom, E. Tolkacheva, S. Uchaikin, J. Wang, A. B. Wilson, Z. Merali, and G. Rose. Thermally assisted quantum annealing of a 16-qubit problem. *Nat. Comm.*, 4:1903, 2013.

[28] R. Douc and O.Cappé. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 64–69. IEEE, 2005.

[29] V. Dupuis, E. Vincent, J.-P. Bouchaud, J. Hammann, A. Ito, and H. A. Katori. Aging, rejuvenation, and memory effects in Ising and Heisenberg spin glasses. *Phys. Rev. B*, 64:174204, 2001.

[30] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

[31] S. Edwards and P. W. Anderson. Theory of spin glasses. *J.Phys.F*, 5:965–974, 1975.

[32] L. A. Fernandez, V. Martin-Mayor, G. Parisi, and B. Seoane. Temperature chaos in 3D Ising spin glasses is driven by rare events. *Europhys. Lett.*, 103(6):67003, 2013.

[33] A. L. Ferreira and M. A. Barroso. Temperature and density extrapolations in canonical ensemble Monte Carlo simulations. *Phys. Rev. E*, 61:1195–1198, 2000.

[34] Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.

[35] D. S. Fisher and D. A. Huse. Ordered phase of short-range Ising spin-glasses. *Phys. Rev. Lett.*, 56:1601, 1986.

[36] D. S. Fisher and D. A. Huse. Absence of many states in realistic spin glasses. *J. Phys. A*, 20:L1005, 1987.

[37] D. S. Fisher and D. A. Huse. Equilibrium behavior of the spin-glass ordered phase. *Phys. Rev. B*, 38:386, 1988.

[38] D. S. Fisher and D. A. Huse. Directed paths in a random potential. *Phys. Rev. B*, 43:10728, 1991.

[39] M. H. Freedman, A. Y. Kitaev, M. J. Larsen, and Z. Wang. Topological Quantum Computation. (arXiv:quant-ph/0101025), 2002.

[40] C. Geyer. Monte Carlo Maximum Likelihood for Dependent Data. In E. M. Keramidas, editor, *Computing Science and Statistics: 23rd Symposium on the Interface*, page 156, Fairfax Station, 1991. Interface Foundation.

[41] A. K. Hartmann and H. Rieger. *Optimization Algorithms in Physics*. Wiley-VCH, Berlin, 2001.

[42] A. K. Hartmann and H. Rieger. *New Optimization Algorithms in Physics*. Wiley-VCH, Berlin, 2004.

[43] Alexander K. Hartmann. Scaling of stiffness energy for three-dimensional $\pm J$ Ising spin glasses. *Phys. Rev. E*, 59:84–87, 1999.

[44] M Hasenbusch. Monte-Carlo simulation with fluctuating boundary-conditions. *Physica A*, 197:423–435, 1993.

[45] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. Lidar. Probing for quantum speedup in spin glass problems with planted solutions. (arXiv:quant-phys/1502.01663), 2015.

[46] J.J. Houdayer, F. Krzakala, and O. C. Martin. Large-scale low-energy excitations in 3-d spin glasses. *Eur. Phys. J. B.*, 18:467, 2000.

[47] K. Hukushima. Domain-wall free energy of spin-glass models: Numerical method and boundary conditions. *Phys. Rev. E*, 60:3606–3613, 1999.

[48] K. Hukushima and Y. Iba. Population annealing and its application to a spin glass. In James E. Gubernatis, editor, *THE MONTE CARLO METHOD IN THE PHYSICAL SCIENCES: Celebrating the 50th Anniversary of the Metropolis Algorithm*, volume 690, pages 200–206. AIP, 2003.

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

[50] Wolfhard Janke. Monte Carlo methods in classical statistical physics. In H. Fehske, R. Schneider, and A. Weie, editors, *Computational Many-Particle Physics*, volume 739 of *Lecture Notes in Physics*, pages 79–140. Springer Berlin Heidelberg, 2008.

[51] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[52] P. E. Jönsson, R. Mathieu, P. Nordblad, H. Yoshino, H. A. Katori, and A. Ito. Nonequilibrium dynamics of spin glasses: Examination of the ghost domain scenario. *Phys. Rev. B*, 70:174402, 2004.

[53] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado. Error Threshold for Color Codes and Random 3-Body Ising Models. *Phys. Rev. Lett.*, 103:090501, 2009.

[54] H. G. Katzgraber, F. Hamze, and R. S. Andrist. Glassy Chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines. (arXiv:1401.1546), 2014.

[55] H. G. Katzgraber, F. Hamze, Z. Zhu, and A. J. Ochoa. Seeking Quantum Speedup Through Spin Glasses: The Good, the Bad, and the Ugly. in preparation, 2015.

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

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

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

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

[60] H. G. Katzgraber and A. P. Young. Monte Carlo simulations of spin-glasses at low temperatures: Effects of free boundary conditions. *Phys. Rev. B*, 65:214402, 2002.

[61] H. G. Katzgraber and A. P. Young. Monte Carlo studies of the one-dimensional Ising spin glass with power-law interactions. *Phys. Rev. B*, 67:134410, 2003.

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

[63] A. Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Ann. Phys.*, 303:2, 2003.

[64] F. Krzakala and O. C. Martin. Spin and link overlaps in 3-dimensional spin glasses. *Phys. Rev. Lett.*, 85:3013, 2000.

[65] F. Krzakala and O. C. Martin. Chaotic temperature dependence in a model of spin glasses. A random-energy random-entropy model. *Eur. Phys. J. B*, 28:199, 2002.

[66] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[67] J. W. Landry and S. N. Coppersmith. Ground states of two-dimensional $\pm J$ Edwards-Anderson spin glasses. *Phys. Rev. B*, 65:134404, 2002.

[68] P. le Doussal. Chaos and residual correlations in pinned disordered systems. (cond-mat/0505679), 2005.

[69] A. Lucas. Ising formulations of many NP problems. *Front. Physics*, 12:5, 2014.

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

[71] J. Machta and R. Ellis. Monte Carlo methods for rough free energy landscapes: Population annealing and parallel tempering. *J. Stat. Phys.*, 144:541–553, 2011.

[72] S. R. McKay, A. N. Berker, and S. Kirkpatrick. Spin-Glass Behavior in Frustrated Ising Models with Chaotic Renormalization-Group Trajectories. *Phys. Rev. Lett.*, 48:767, 1982.

[73] W. L. McMillan. Domain-wall renormalization-group study of the three-dimensional random Ising model. *Phys. Rev. B*, 30:R476, 1984.

[74] W. L. McMillan. Domain-wall renormalization-group study of the two-dimensional random Ising model. *Phys. Rev. B*, 29:4026, 1984.

[75] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.

[76] M.Ferrario, G.Ciccotti, and K.Binder, editors. *LECTURE NOTES IN PHYSICS Computer Simulations in Condensed Matter: From Materials to Chemical Biology-Vol.2*. Springer, 2006.

[77] A. A. Middleton. Improved extremal optimization for the Ising spin glass. *Phys. Rev. E*, 69:055701, 2004.

[78] A. Alan Middleton. Extracting thermodynamic behavior of spin glasses from the overlap function. *Phys. Rev. B*, 87:220201, Jun 2013.

[79] Cécile Monthus and Thomas Garel. Typical versus averaged overlap distribution in spin glasses: Evidence for droplet scaling theory. *Phys. Rev. B*, 88:134204, Oct 2013.

[80] C. Moore and S. Mertens. *The Nature of Computation*. Oxford University Press, 2011.

[81] J. J. Moreno, H. G. Katzgraber, and A. K. Hartmann. Finding low-temperature states with parallel tempering, simulated annealing and simple monte carlo. *International Journal of Modern Physics C*, 14(3):285 – 302, 2003.

[82] C. M. Newman and D. L. Stein. Multiple states and thermodynamic limits in short-ranged Ising spin-glass models. *Phys. Rev. B*, 46:973, 1992.

[83] C. M. Newman and D. L. Stein. Non-mean-field behavior of realistic spin glasses. *Phys. Rev. Lett.*, 76:515–518, 1996.

[84] C. M. Newman and D. L. Stein. Simplicity of state and overlap structure in finite-volume realistic spin glasses. *Phys. Rev. E*, 57:1356–1366, 1998.

[85] C M Newman and D L Stein. The state(s) of replica symmetry breaking: Mean field theories vs short-ranged spin glasses. *J. Stat. Phys.*, 106:213–244, 2002.

[86] C. M. Newman and D. L. Stein. TOPICAL REVIEW: Ordering and broken symmetry in short-ranged spin glasses. *J. Phys.: Condensed Matter*, 15:1319, 2003.

[87] D. Nigg, M. Mueller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt. Experimental Quantum Computations on a Topologically Encoded Qubit. (arxiv:quant-phys/1403.5426), 2014.

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

[89] P. Nordblad and P. Svendlidh. Experiments on spin glasses. In A. P. Young, editor, *Spin glasses and random fields*. World Scientific, Singapore, 1998.

[90] M. Palassini and A. P. Young. Triviality of the ground state structure in Ising spin glasses. *Phys. Rev. Lett.*, 83(24):5126 – 5129, 1999.

[91] M. Palassini and A. P. Young. Nature of the spin glass state. *Phys. Rev. Lett.*, 85:3017–3020, 2000.

[92] Dmitry Panchenko. The Sherrington-Kirkpatrick model: An overview. *J. Stat. Phys.*, 149(2):362–383, 2012.

[93] G. Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754, 1979.

[94] G. Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *J. Phys. A*, 13:L115, 1980.

[95] G. Parisi. The order parameter for spin glasses: a function on the interval 0–1. *J. Phys. A*, 13:1101, 1980.

[96] G. Parisi. Order parameter for spin-glasses. *Phys. Rev. Lett.*, 50:1946, 1983.

[97] G. Parisi. Spin glasses and replicas. *Physica A*, 124:523, 1984.

[98] G. Parisi. Some considerations of finite dimensional spin glasses. *J. Phys. A*, 41:324002, 2008.

[99] Kristen L Pudenz, Tameem Albash, and Daniel A Lidar. Error-corrected quantum annealing with hundreds of qubits. *Nat Commun*, 5, 2014.

[100] R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Rev. Mod. Phys.*, 58(3):765, 1986.

[101] N. Read. Short-range Ising spin glasses: the metastate interpretation of replica symmetry breaking. (arXiv:cond-mat/1407.4136), 2014.

[102] T. Rizzo and A. Crisanti. Chaos in Temperature in the Sherrington-Kirkpatrick Model. *Phys. Rev. Lett.*, 90:137201, 2003.

[103] F. Romá, S. Risau-Gusman, A. J. Ramirez-Pastor, F. Nieto, and E. E. Vogel. The ground state energy of the edwards-anderson spin glass model with a parallel tempering Monte Carlo algorithm. *Physica A: Statistical Mechanics and its Applications*, 388(14):2821–2838, 2009.

[104] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer. Defining and detecting quantum speedup. *Science*, 345(420), 2014.

[105] Juan J. Ruiz-Lorenzo. Low Temperature Properties of Finite Dimensional Ising Spin Glasses: (some) Numerical simulations. *ArXiv e-prints*, 2003.

[106] M. Sales and H. Yoshino. Fragility of the free-energy landscape of a directed polymer in random media. *Phys. Rev. E*, 65:066131, 2002.

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

[108] M. Sasaki, K. Hukushima, H. Yoshino, and H. Takayama. Scaling Analysis of Domain-Wall Free Energy in the Edwards-Anderson Ising Spin Glass in a Magnetic Field. *Phys. Rev. Lett.*, 99:137202, 2007.

[109] T. Schilling and F. Schmid. Computing absolute free energies of disordered structures by molecular simulation. *The Journal of Chemical Physics*, 131(23):231102, 2009.

[110] D. Sherrington and S. Kirkpatrick. Solvable model of a spin glass. *Phys. Rev. Lett.*, 35:1792–1796, 1975.

[111] Caterina De Simone, Martin Diehl, Michael Jünger, Petra Mutzel, Gerhard Reinelt, and Giovanni Rinaldi. Exact ground states of Ising spin glasses: New experimental results with a branch and cut algorithm. *J. Stat. Phys.*, 80:487–496, 1995.

[112] R. D. Somma, D. Nagaj, and M. Kieferová. Quantum Speedup by Quantum Annealing. *Phys. Rev. Lett.*, 109:050501, 2012.

[113] D.L. Stein and C.M. Newman. *Spin Glasses and Complexity*. Primers in Complex Systems. Princeton University Press, 2013.

[114] R. H. Swendsen and J.-S. Wang. Replica Monte Carlo simulations of spin glasses. *Phys. Rev. Lett.*, 57:2607–2609, 1986.

[115] C. K. Thomas, D. A. Huse, and A. A. Middleton. Chaos and universality in two-dimensional Ising spin glasses. *Phys. Rev. Lett.*, 107:047203, 2011.

[116] C. K. Thomas and A. A. Middleton. Matching Kasteleyn cities for spin glass ground states. *Phys. Rev. B*, 76:220406(R), 2007.

[117] Creighton K. Thomas, David A.Huse, and A.Alan Middleton. Chaos and universality in two-dimensional Ising spin glasses. *Phys. Rev. Lett.*, 107:047203, 2011.

[118] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[119] S. Trebst and U.H.E. Hansmann. Optimized folding simulations of protein A. *The European Physical Journal E*, 24(3):311–316, 2007.

[120] Simon Trebst, Matthias Troyer, and Ulrich H.E. Hansmann. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.*, 124:174903, 2006.

[121] F. Wang and D. P. Landau. An efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050, 2001.

[122] W. Wang. Simulating thermal boundary conditions of spin-lattice models using parallel tempering Monte Carlo. *arXiv:1507.08686*.

[123] W. Wang. Measuring free energy in spin-lattice models using parallel tempering Monte Carlo. *Phys. Rev. E*, 91:053303, 2015.

[124] W. Wang, J. Machta, and H. G. Katzgraber. Chaos in Spin Glasses Revealed Through Thermal Boundary Conditions. *arXiv:1505.06222*.

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

[126] W. Wang, J. Machta, and H. G. Katzgraber. 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.

[127] W. Wang, J. Machta, and H. G. Katzgraber. Population annealing for large scale spin glass simulations. (2015), in preparation.

[128] L. E. Wenger and P. H. Keesom. Calorimetric investigation of a spin-glass alloy: CuMn. *Phys. Rev. B*, 13:4053–4059, 1976.

[129] A. P. Young, editor. *Spin Glasses and Random Fields*. World Scientific, Singapore, 1998.

[130] B. Yucesoy. *Replica Exchange Monte Carlo Simulations of the Ising Spin Glass: Static and Dynamic Properties*. PhD thesis, University of Massachusetts-Amherst, 2013.

[131] B. Yucesoy, H. Katzgraber, and J. Machta. Evidence of Non-Mean-Field-Like Low-Temperature Behavior in the Edwards-Anderson Spin-Glass Model. *Phys. Rev. Lett.*, 109(17):177204, 2012.

[132] B. Yucesoy, H. G. Katzgraber, and J. Machta. Yucesoy, Katzgraber, and Machta reply:. *Phys. Rev. Lett.*, 110:219702, 2013.

[133] B. Yucesoy, Helmut G. Katzgraber, and J. Machta. Evidence of Non-Mean-Field-Like Low-Temperature Behavior in the Edwards-Anderson Spin-Glass Model. *Phys. Rev. Lett.*, 109:177204, 2012.

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

[135] Z. Zhu, A. J. Ochoa, F. Hamze, S. Schnabel, and H. G. Katzgraber. Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities. in preparation, 2015.

[136] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber. Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension. 2015. (cond-mat/1501.05630).

[137] R. W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics*, 22(8), 1954.