2010

Efficiency of Parallel Tempering for Ising Systems

Stephan Burkhardt
University of Massachusetts Amherst

EFFICIENCY OF PARALLEL TEMPERING FOR ISING SYSTEMS

A Thesis Presented

by

STEPHAN BURKHARDT

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

MASTER OF SCIENCE

September 2010

Department of Physics

EFFICIENCY OF PARALLEL TEMPERING FOR ISING SYSTEMS

A Thesis Presented

by

STEPHAN BURKHARDT

Approved as to style and content by:

Jonathan Machta, Co-chair

Christian Santangelo, Co-chair

Donald Candela, Department Chair
Department of Physics

ACKNOWLEDGMENTS

First, I would like thank my advisor Prof. Jonathan Machta who did not only inspire and encourage me to pursue this research, but whose help and insight into statistical physics also assisted in completing this thesis.

I would like to moreover express my profound gratitude to my girlfriend Rebecca for her patience, encouragement and understanding during my whole stay at Umass Amherst. Finally, my thanks also go to my parents who generously supported me financially and also supported me in my choice to pursue a masters degree during my stay in the United States.

ABSTRACT

EFFICIENCY OF PARALLEL TEMPERING FOR ISING SYSTEMS

SEPTEMBER 2010

STEPHAN BURKHARDT

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jonathan Machta and Professor Christian Santangelo

The efficiency of parallel tempering Monte Carlo is studied for a two-dimensional Ising system of length L with $N = L^2$ spins. An external field is used to introduce a difference in free energy between the two low temperature states.

It is found that the number of replicas R_{opt} that optimizes the parallel tempering algorithm scales as the square root of the system size N. For two symmetric low temperature states, the time needed for equilibration is observed to grow as $L^{2.18}$. If a significant difference in free energy is present between the two states, this changes to $L^{1.02}$.

It is therefore established that parallel tempering is sped up by a factor of roughly L if an asymmetry is introduced between the low temperature states. This confirms previously made predictions for the efficiency of parallel tempering. These findings should be especially relevant when using parallel tempering for systems like spin glasses, where no information about the degeneracy of low temperature states is available prior to the simulation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Most of our understanding about the world is based on models. In fields like particle physics, we try to describe the world as accurately as possible with these models, including every possible aspect of reality in them. In other fields like statistical physics, we try to keep these models as simple as possible, including only the aspects we believe to be crucial for describing the behavior of a system. But even these simplified models can not always be completely understood by purely analytical means. Today, more and more fields of science therefore make use of numerical methods in order analyze these models. For stochastic models, one of the most widespread class of algorithms are Monte Carlo methods, especially in the field of statistical physics. Monte Carlo methods are a powerful class of algorithms that is not only used in physics, but also in fields such as chemistry, biology, material science, finance and statistics.

Especially in biology, but also in condensed matter physics, many problems are characterized by conflicting forces on a microscopic level, forbidding simultaneous minimization. This effect is commonly known as frustration and usually makes systems very hard to simulate using traditional Monte Carlo methods. Today, the standard method fur such systems is parallel tempering, also known as replica exchange Monte Carlo [9, 18].

The general idea of parallel tempering is to simulate multiple copies of a system, each copy at a different temperature. While the copies at a low temperature accurately sample a local region of the configuration space, they also tend to get stuck

in local minima. Copies at a higher temperature on the other hand sample a larger part of the configuration space, but by doing so they can not give accurate information about the regions we are actually interested in. Parallel tempering allows both these behaviors to cooperate and is therefore very effective for systems where the presence of many local minima otherwise hampers the simulation of the system. But the efficiency of parallel tempering depends on the properties of the system at all the simulated temperatures and it is therefore often not easy to predict, how parallel tempering will perform for a specific system.

One class of systems, where the question about the efficiency of parallel tempering is highly relevant are spin-glasses. Spin glasses are systems with randomized interactions terms and are strongly affected by frustration. As of today, parallel tempering is the only method known to be effective for simulating spin glasses and is therefore used extensively for analyzing them. One question intensely studied in the last years concerns the low temperature states of three-dimensional spin glasses. There are two competing theories for describing 3-dimensional spin glasses in the thermodynamic limit: the droplet [3] and the replica symmetry breaking picture[16]. Both theories make very different predictions when it comes to the number of pure thermodynamic states at low temperatures. For this reason, evidence supporting or contradicting either of the models is most likely to be found in precision simulations of low temperature spin glasses. But these measurements might exhibit a significant bias if the efficiency of parallel tempering is very different depending on whether there are nearly degenerate free energy minima.

This thesis will deal with the efficiency of parallel tempering for the two-dimensional Ising model, a system much simpler than spin glasses. Although the Ising model in two dimensions is the simplest system with a symmetry breaking phase transition, the results should nevertheless be generalizable to more complicated systems. Normally, the Ising model possesses a discrete symmetry in the low temperature phase, leading

to degenerate ground states. But this symmetry can be lifted by introducing an external field. The case of an Ising model without and with an external field can then be seen as similar to e.g. a spin glass with a degenerate or non-degenerate free energy minima. Despite the simplicity of the Ising model, the results of this thesis should therefore prove interesting for better understanding the parallel tempering algorithm even for more complex systems.

CHAPTER 2

THE ISING MODEL AND MONTE CARLO METHODS

2.1 The Ising Model

The Ising model is easily the most studied and best understood model of the whole field of statistical physics. Introduced in 1920 by Wilhelm Lenz and first studied in 1925 by Ernst Ising[8], the Ising model is a model of ferromagnetism. It consists of a number of spins s_i that points either up ($s_i = +1$) or down ($s_i = -1$) and are arranged on a regular lattice of dimension n. Each of the spins only interacts with its $2n$ direct neighbors and the energy of the system is given by the Hamiltonian

$$E = -J \sum_{\langle i,j \rangle} s_i s_j \tag{2.1}$$

where the sum over $\langle i, j \rangle$ denotes a sum over all neighboring pairs of spins.

Although at first, the model was thought to not exhibit phase transitions in any dimension, it quickly turned out that it shows a broken symmetry phase for two or more dimensions [4]. Despite its simplicity it is therefore an interesting model for understanding the behavior in the vicinity of critical points. The phase transition is between a high temperature phase where all spins are pointing more or less in random directions and an ordered low temperature phase where almost all spins of the lattice have the same alignment. When looking at (2.1) we can see that the model is symmetric to flipping all the spins at the same time($s'_i = -s_i$). In the ordered phase the system can therefore be either in a state where almost all spins are $+1$ or where almost all of them are -1.

It is possible to introduce an external field H to the Ising model by using the following Hamiltonian:

$$E = -J \sum_{\langle i,j \rangle} s_i s_j. - H \sum_i s_i. \tag{2.2}$$

This introduces a preferred direction and spins will be more inclined to have the orientation +1.

2.2 Monte Carlo Methods

Monte Carlo algorithms are usually defined as a class of algorithms that use random sampling to solve a problem. Although the first Monte Carlo algorithms date back to the middle of the nineteenth century, they only gained their modern name and widespread acceptance with the advent of electronic computers [14]. Today, Monte Carlo methods are used in a wide variety of fields, including chemistry, biology, material science, finance, statistics and of course physics. The general idea of every Monte Carlo algorithm is to solve an integral by sampling random points of the integrand.

To understand how a Monte Carlo algorithm works, it is useful to remember that almost every problem in statistical mechanics can be expressed as calculating the expectation value of a quantity A for a certain ensemble. Calculating this value can be expressed as a sum (or for the continuous case an integral) of the form:

$$\begin{aligned} \langle A \rangle &= \sum_{s \in \Omega} A(s) P(s) \\ &= \sum_{s \in \Omega} A(s) \left(\frac{w(s)}{Z} \right) \end{aligned} \tag{2.3}$$

Where Ω is the set of all possible states (also called configurations) of the system, $w(s)$ is the relative probability that the system is in this particular state and

$Z = \sum_{s \in \Omega} w(s)$ is the normalization constant for the probabilities. The most common example is a thermodynamic equilibrium at temperature T where each state is weighted with its Boltzmann probability:

$$w(s) = \exp\left(-\beta E_s\right) \tag{2.4}$$

where E_s is the energy of the state s and $\beta = \frac{1}{k_B T}$ the Boltzmann factor. In this case Z is the partition function:

$$Z = \sum_{s \in \Omega} \exp\left(-\beta E_s\right) \tag{2.5}$$

Unless stated otherwise, every system mentioned from here on should be assumed to be in thermodynamic equilibrium.

While the analytical evaluation of the sum in (2.3) has been done for some cases [15, 1], we need to rely on numerical methods for most of the systems we are interested in today. But even exact numerical evaluations of the sum are usually not possible because the space Ω of all possible states grows exponentially with the size of the problem. Monte Carlo methods come into play when the exact evaluation of (2.3) is not possible, but the weights $w(s)$ can still be calculated. This is true for almost all systems in thermodynamic equilibrium, as here the weight only depends on the energy of the state. It is important to note that most Monte Carlo methods do not need access to the normalization factor Z. Deriving an analytical expression for the partition function Z is usually extremely difficult and often even tantamount to solving the system.

The name "Monte Carlo Method" is used to describe algorithms that solve integrals through the use of random numbers. The most straightforward Monte Carlo

algorithm is to simply pick M random configurations $\{s_k | s_k \in \Omega\}$ and use the weights $w(s_k)$ to calculate the weighted mean for the desired quantity A:

$$\bar{A} = \frac{\sum_{k=0}^{M} A(s_k) w(s_k)}{\sum_{k=0}^{M} w(s_k)}$$

By the law of large numbers we expect that the mean $\bar{A}$ calculated in this way is an estimate for $\langle A \rangle$ and will therefore approach the real expectation value $\langle A \rangle$ if M is large enough. This can be seen as interpreting (2.3) in a different way: calculating a value $A'(s) = \frac{A(s)w(s)}{Z}$ over an ensemble where $P(s) = \text{const}$.

The problem that arises when using this approach is that the space of all possible configurations is usually big and of a high dimension, thus making it impossible to sample a significant fraction of it in any reasonable amount of time. When studying systems in thermodynamic equilibrium, this problem is exacerbated by the fact that usually only a small fraction of all states have a high weight $w(s)$. The weight of the other states is so small that these states hardly influence the expectation value $\langle A \rangle$. Only by sampling the states with a high weight $w(s)$ can we improve the estimate $\bar{A}$, but by picking random configurations the algorithm spends almost all of its time sampling low weighted states without any significance for $\langle A \rangle$. Therefore all Monte Carlo methods applicable to systems in thermodynamic equilibrium will try to only sample the part of the configuration space that has a high weight $w(s)$ and therefore plays an important role for $\langle A \rangle$. This concept is known as *importance sampling*.

In fact this is what happens in real thermodynamical systems as well. Many configurations are so unlikely, that they will never occur within any reasonable time frame. The time one litre of gas at atmospheric pressure and room temperature needs to go through all possible configurations is for example $10^{10^{23}}$ times the lifetime of our universe (see [14], chapter 2.2). For this reason it makes sense to concentrate only on the configurations with a significant weight.

2.2.1 The Metropolis Algorithm

Almost all Monte Carlo algorithms in use today are Markov Chain Monte Carlo algorithms. This means that instead of generating a number of independent configurations, a Markov process is used to generate a chain of configurations $\{s_k \ | k \in \{1, ..., M\}\}$ where

$$P(s_k = s') = \sum_{s \in \Omega} P(s_{k-1} = s) P(s \to s')$$

The underlying idea is that if we already found a configurations with high weight w(s), similar configurations s' will have a similarly high weight.

It can be shown ([14], chapter 2.2.3) that such a Markov process will correctly sample a distribution $P(s)$ in the limit of $M \to \infty$ if two requirements are met:

- The process has to be ergodic, meaning that there has to be a finite probability to get from any configuration μ to any other configuration ν in a finite amount of steps. This is a necessary condition for sampling the distribution.

- The detailed balance equation is satisfied

$$\frac{P(\mu)}{P(\nu)} = \frac{P(\mu \to \nu)}{P(\nu \to \mu)} \tag{2.6}$$

  This condition is not necessary, but together with ergodicity it is sufficient for guaranteeing that the Markov process samples the correct distribution. It should be noted that we don't have to care about normalization factors in $P(\mu)$, as they cancel out.

Even though we are guaranteed to sample the distribution correctly in the limit case of $M \to \infty$, there will usually be a strong correlation between two configurations s_k and s_{k+i} that are i steps apart. This correlation means that we will need to wait a certain amount of steps before two configurations s_t and s_{t+i} are effectively uncorrelated. How we can estimate the time needed to generate to independent samples is explained in chapter 2.3.

The first Monte Carlo algorithm applied to physics was a Markov Chain Monte Carlo algorithm called the Metropolis algorithm [13]. As it's formulation is quite general, many of the Monte Carlo algorithms used today can be seen as special cases of the Metropolis algorithm. For generating a new configuration μ_{n+1} from the configuration μ_n, the algorithm uses the following two steps ([14], chapter 3.1):

- Propose a new state ν according to a selection probability $g(\mu_n \to \nu)$.
- Accept this transition to the new state ν with a probability $A(\mu_n \to \nu)$. If the transition is accepted, μ_{n+1} is ν, otherwise $\mu_{n+1} = \mu_n$.

The resulting Markov chain will correctly sample a distribution $P(\mu)$ if the criteria of ergodicity and detailed balance are met. If we assume all acceptance probabilities $A(\mu \to \nu)$ to be non-zero, ergodicity is only a demand on the selection probabilities $g(\mu \to \nu)$, that can be quite easily satisfied for most simple algorithms. The more complicated the algorithm, the harder it is often to guarantee ergodicity, but for all algorithms presented in this theses, ergodicity will not be hard to proof. For detailed balance, the probabilities need to satisfy equation (2.6):

$$\frac{P(\nu)}{P(\mu)} = \frac{g(\mu \to \nu)A(\mu \to \nu)}{g(\nu \to \mu)A(\nu \to \mu)} \tag{2.7}$$

The selection probabilities are often chosen symmetric, so that $g(\mu \to \nu) = g(\nu \to \mu)$. This means that they cancel out of (2.7) and therefore the acceptance probabilities have to satisfy the simplified detailed balance equation:

$$\frac{P(\nu)}{P(\mu)} = \frac{A(\mu \to \nu)}{A(\nu \to \mu)} \tag{2.8}$$

This can be easily fulfilled by choosing the following acceptance probability (note that the normalization constants of $P(s)$ cancel out);

$$A(\mu \to \nu) = \min\left[1, \frac{P(\nu)}{P(\mu)}\right] \tag{2.9}$$

For a system in thermodynamic equilibrium this would mean that

$$A(\mu \to \nu) = \min\left[1, \exp\left\{-\beta(E_\nu - E_\mu)\right\}\right] \tag{2.10}$$

In order to make the sampling of the probability distribution efficient, the acceptance probabilities as given by (2.10) should not be to small, so that the algorithm does not get "stuck" in one state. This means that we should try to propose new states with a similar energy to the current one. The easiest way to guarantee this is to only propose small changes to the system. But on the other hand this makes the algorithm inefficient, because the new state will be very similar to the old one and actually sampling the significant part of the probability distribution will take a long time. The significance of this dilemma will be further explained in 2.3.

An interesting note is that the way the algorithm works often imitates what actually happens in real physical processes. Most physical processes do not sample the whole configuration space randomly but are instead driven by small local changes, similar to the steps of the Metropolis algorithm.

2.2.2 A simple Metropolis Algorithm for the Ising Model

A straightforward example of an Metropolis algorithm can be constructed for the Ising model. As the configuration of the Ising Model consists of N spins that can assume values of either $+1$ or -1, an easy way to propose a new configuration ν is to flip a single spin of the old configuration μ. The acceptance probability for this move is given by (2.10). As every spin only interacts with its nearest neighbors, the energy difference between the old and the new configuration can be easily computed. For flipping the spin i it is:

$$\Delta E(i) = 2Hs_i + 2Js_i \sum_{j \in \rho(i)} s_j. \tag{2.11}$$

where s_i is value before flipping the spin i and $\rho(i)$ are all nearest neighbors of the spin i. Using (2.10) the acceptance ratio would then be[1]:

$$A(i) = \min\left[1, \exp\left\{-\beta\, \Delta E(i)\right\}\right] \tag{2.12}$$

As the system consists of $N \gg 1$ spins and $N - 1$ of them stay completely untouched, flipping one spin will not produce a completely new configuration. We will therefore group N proposed flips together and call them a Monte Carlo sweep (MC-sweep). The reason for this is that in order to produce an independent configuration, we need to propose every spin at least once for a flip. One MC-sweep is therefore the shortest possible time needed to create an independent configuration.

The Metropolis algorithm for the Ising model thus has the following form:

1. Pick a random spin i and calculate $\Delta E(i)$.

2. Flip the spin i with a probability of $A(i)$.

3. Repeat the steps 1 and 2 N times. Then take the desired measurement

Even for small systems, one MC-sweep will not be enough to produce a truly independent configuration. But the larger the system size, the less effective will a single MC-sweep be, especially in vicinity of the critical temperature. There will therefore usually be a strong correlation between subsequent measurements. This correlation can be quantified using the autocorrelation time described in section 2.3

[1]As $\Delta E(i)$ can only have 10 different values, it makes sense to calculate all possible value of $\exp\{-\beta\Delta E(i)\}$ and store them in memory, so that the computationally costly calculation of the exponential function only has to be done once at the start of the algorithm.

2.3 Autocorrelation time and problems with free energy barriers

To understand the performance of Monte Carlo algorithms a bit better, it is convenient to consider the free energy F. The importance of the free energy (henceforth abbreviated FE) can be understood by looking at the probability of the system to be in a certain macrostate m which can be realized by $|\Omega_m|$ microstates. In thermodynamic equilibrium, the probability to find the system in the macrostate m is:

$$P(m) = \frac{\sum_{s \in m} \exp\{-\beta E_s\}}{Z} \tag{2.13}$$

$$= \frac{Z_m}{Z} \propto Z_m \tag{2.14}$$

Where Z_m is the partition sum of the ensemble that is restricted to the macrostate m. The partition sum can be expressed in terms of the free energy F by the following equation (see chapter 3.3 of [22]):

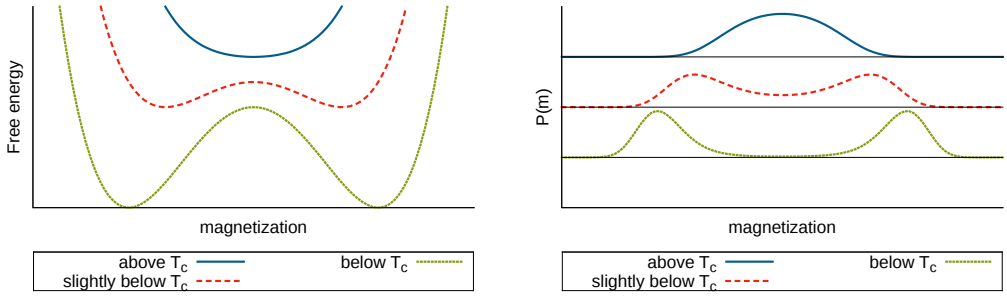
$$F = -kT \log(Z) \tag{2.15}$$

$$\Rightarrow Z = e^{-\beta F} \tag{2.16}$$

This is also true for the partition sum Z_m and the free energy F_m of the macrostate m, leading to:

$$P(m) \sim e^{-\beta F_m} \tag{2.17}$$

This makes it possible to look at a low-dimensional projection of the configuration space and still understand the dynamics of the system. To do this we look at a



(a) Free energy at magnetization (b) Probability of magnetization

Figure 2.1: Sketch of the properties of an Ising system. Below the critical temperature the distribution is bimodal and the two ground states are separated by a barrier in free energy

variable $x(\vec{s})$ that is a function of the state $\vec{s}$ and then identify the ensemble $\hat{X}$ as all states $\vec{s}$ where $x(\vec{s}) = X$:

$$\hat{X} \hat{=} \{\vec{s} \,|\, x(\vec{s}) = X\} = x^{-1}(X)$$

For an Ising system with N spins we could for example choose the magnetization $m = \sum_i \frac{s_i}{N}$, reducing the dimension to one. We can then calculate the free energy as a function of the magnetization and therefore determine the probability of the system being in a state with magnetization m. As every Metropolis MC step can only change the magnetization by $\frac{2}{N}$, the algorithm can only move in small steps (see figure 2.2 for an illustration). The probability for a step that changes the magnetization can be calculated using in an analogue to (2.10), where the energy is replaced by the free energy:

$$P(m \to m') = \min\left[1, \exp\left\{-\beta(F_{m'} - F_m)\right\}\right] \tag{2.18}$$

This means that the standard Metropolis algorithm will most likely follow the slope of the free energy function downhill. Changes that increase the free energy are exponentially suppressed.

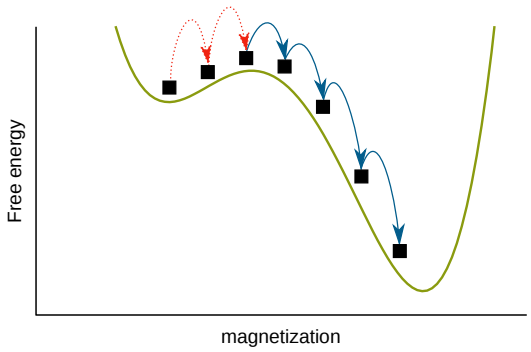
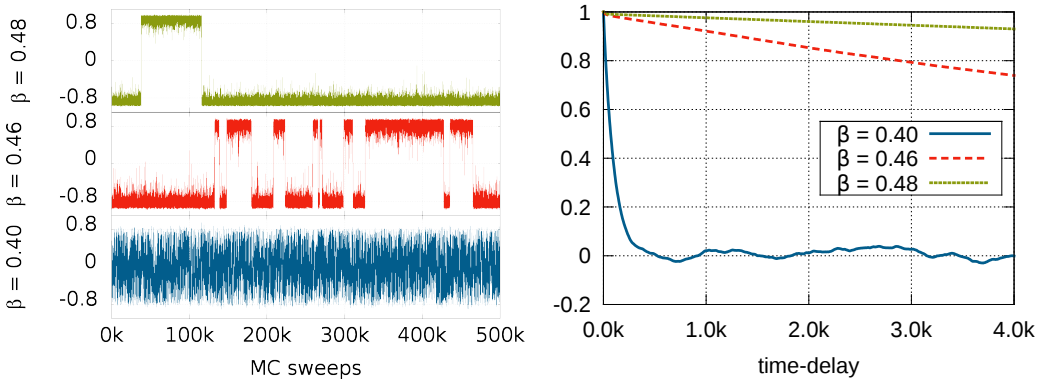


Figure 2.2: Monte Carlo steps that increase the free energy (red, dotted) are suppressed while steps that decrease the free energy (blue, solid) are preferred

We can see in figure 2.1b that the Ising system below the critical temperature has degenerate ground states, due to the symmetry of the problem. Both ground states are separated by a barrier in free energy as can be seen in figure 2.1a. The standard Metropolis algorithm will have problems crossing this barrier because this means traversing the states around $m = 0$ which are high in free energy and therefore very unlikely to be reached by the algorithm[2]. Being an ergodic statistical process, the Markov process will eventually cross the barrier if we only wait long enough, but the mean time needed to do so becomes arbitrarily large as the height of the barrier grows. In figure 2.3a we can see that the barrier crossing becomes less frequent with decreasing temperature. In section 2.2.1 two sufficient conditions are given that guarantee that Markov process actually samples the target distribution: detailed balance and ergodicity. Whereas the detailed balance equation is not a necessary condition and can be replaced by a less strict but necessary equation, ergodicity is a strictly necessary condition to ensure sampling the target distribution (see chapter

[2]The phenomenon is of course well-known for physical systems. For example, hysteresis leaves a ferromagnetic metal magnetized even after any external field has been removed. Although all magnetization directions should be equally likely, the metal will not change or lose its magnetization as long as the temperature stays below some critical point. This does not mean however, that the effect is desirable in Monte Carlo simulations where we are interested in thermodynamic equilibrium properties.



(a) Magnetization (b) Autocorrelation function

Figure 2.3: Simulation for a 2D Ising lattice. The first figure shows the magnetization as function of time, the second the time-delayed autocorrelation functions for the magnetizations. The simulation was done for three different temperatures roughly corresponding to the ones used in figure 2.1a

2.2 of [14]). Even a system with multiple minima in free energy, that are separated by high free energy barriers, is ergodic in theory. But if the average time needed to cross the barrier is equal or bigger than the runtime of the simulation, ergodicity is effectively broken. The simulation gets stuck in a local minimum of free energy and samples only this part of the configuration space instead of the whole target distribution. This becomes especially problematic when the simulation never reaches the ground state because a barrier in free energy keeps it from leaving a local minimum (see for example figure 2.2).

A good indicator to quantify the time needed to cross the barriers and accurately sample the whole system is the autocorrelation time. In order to compute it, one first needs to calculate the time-displaced autocorrelation function $\chi(t)$ of some property x of the model:

$$\chi(t) = \frac{1}{T} \frac{\int_0^T dt' \left[x(t') - \langle x \rangle\right] \left[x(t'+t) - \langle x \rangle\right]}{\mathrm{Var}(x)} \tag{2.19}$$

Where $x(t) = x(\vec{s}(t))$ is the property x of the system at the time t of the simulation, $\langle x \rangle$ the mean value and $\mathrm{Var}(x)$ the variance of the property x.

The property x needs to be chosen in a way that lifts the degeneracy of different minima in the free energy function, otherwise it will not be able to pick up the differences between them. For the Ising system, a good choice would for example be the magnetization m as defined above.

The autocorrelation function tells us how strongly x is correlated over time, or how good we can predict $x(t'+t)$ from $x(t')$. In figure 2.3b we can see that the system has a quickly decaying autocorrelation function if it is simulated above the critical temperature. But if simulated below the critical temperature, the autocorrelation time decays much slower. This corresponds to the problems the simulation has at equilibrating the system in presence of a free energy barrier.

As we normalize with the variance, $\chi(0)$ always equals one. For increasing t it eventually undergoes an exponential decay:

$$\chi(t) \sim e^{-t/\tau} \tag{2.20}$$

There are two different ways to extract a timescale from the time-delayed autocorrelation function. The first method is to integrate over $\chi(t)$ to receive the so called integrated autocorrelation time:

$$\tau_{\mathrm{int}} = \int_0^\infty \chi(t) \approx \int_0^\infty e^{-t/\tau} = \tau. \tag{2.21}$$

The integrated autocorrelation time gives us the typical timescale needed by the Markov process to generate an independent configuration. In other words, if we want to take n independent measurements of the system, we need to run the algorithm for the time τn.

The second way to extract a timescale is through fitting exponential tail of $\chi(t)$ with an exponential function $C * \exp(-t/\tau)$. The parameter τ used for the fit is then called the exponential autocorrelation time τ_{exp}. The exponential autocorrelation time is related to the time the simulation needs to reach the equilibrium distribution and is therefore a good measure for the efficiency of the algorithm. If τ_{exp} is too large, we need to run an algorithm for a very long time before measurements give reliable information about the system.

From equation (2.21) it might appear that τ_{exp} and τ_{int} are identical. This is however not the case, as the approximation in (2.21) does usually not hold. If the autocorrelation function indeed is an exponential, bot values would be identical, but this is usually not the case. Instead the exponential autocorrelation time usually describes the approach to equilibrium, whereas the integrated autocorrelation time describes the efficiency of sampling in equilibrium.

At first it might seem that the exponential autocorrelation time would strongly depend on the property x that is used to compute the autocorrelation function. This is, however, not true. The reason for this is that it is possible to relate the autocorrelation time to the eigenvalues of the transition matrix of the Markov chain. Showing this is beyond the scope of this thesis (for details, see chapter 3.3.2 in [14]). The important fact following from this though is that the ability to measure the autocorrelation time does not strictly depend on the property x we chose for computing the autocorrelation function. The only condition is that x takes different values for different ground states and does therefore not hide the existence of the different states.

In the presence of significant barriers in the free energy landscape equilibration times for the standard Metropolis algorithm will grow exponentially with the height of said barriers. This means that even relatively simple systems such as a 2D Ising system below the critical temperature will become impossible to simulate due to the long time needed for even extracting a handful of independent measurements. This

is especially true for systems with very rough free energy landscapes, such as protein folding or spin glasses.

As we are mostly interested in how long the algorithm needs to find the ground state and the equilibrium distribution of a system, we will be using the exponential autocorrelation time in this thesis. From here on, "autocorrelation time" will therefore denote the exponential autocorrelation time calculated for the time delayed autocorrelation function. When dealing with an Ising model, this time delayed autocorrelation function will be calculated from the magnetization.

2.4 Parallel Tempering

As we have seen in chapter 2.3, most Markov Chain Monte Carlo methods have problems sampling rough free energy landscapes. Although algorithms that overcome this problem do exist for special cases (see for example [21] and [20]), it would be preferable to have a class of algorithms that is applicable to a wider range of systems. One such algorithm is parallel tempering [7, 19, 5]. Parallel tempering (PT) uses the fact that the shape of the free energy landscape varies greatly with the temperature of the system. What might be a very rough FE-landscape at low temperatures is usually quite smooth at high temperatures. In terms of autocorrelation times this means that τ is high for low temperatures and low for high temperatures. The PT algorithm makes use of this to allow crossing FE barriers at high temperatures and sampling the desired statistics at a much lower target temperature.

This facilitation of barrier crossing is achieved by simulating multiple replicas of the same system in parallel using a standard Monte Carlo algorithm, all of them at a different temperature. These replicas are simulated independently of each other most of the time, but every so often we try to exchange the state of two replicas. For the R replicas we use a set of temperatures $T_1, T_2, \cdots, T_R$, where T_1 is the highest temperature and the lowest temperature T_R, also called the target temperature, is

the one where we want to actually study the system. States can move up to a replica with a higher temperature, cross a barrier there and then move down to the target temperature T_R again.

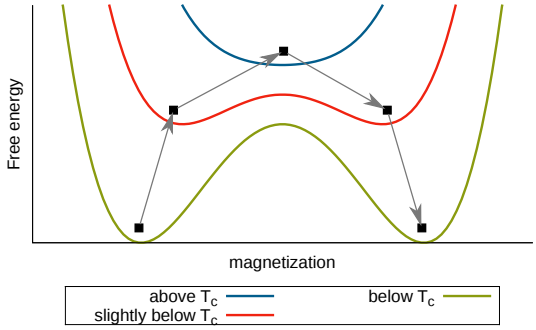


Figure 2.4: Example of how parallel tempering can work around barriers in free energy. The system does not directly cross the barrier but instead changes to a temperature where the barrier does not exist.

The idea is that although the standard Monte Carlo algorithm at T_R could sample the correct statistic of the system, it spends most of the time stuck in local minima of the FE-landscape. The exchange mechanism "feeds" it with new states located in other minima. If the number of replicas R and the set of temperature $T_1, T_2, \cdots, T_R$ are chosen appropriately, the autocorrelation time can be many orders of magnitude smaller than the one of a normal Metropolis algorithm.

In the implementation used here, every replica will be equilibrated using a fixed number of standard Metropolis MC steps. Then R random neighbouring pairs of replicas are proposed for an exchange of states, after which the individual replicas are again equilibrated etc. Each of these combination of steps of the individual replicase and proposed exchanges will be called a PT sweep.

To understand why this algorithm works and how it satisfies detailed balance and ergodicity, let us consider it as a form of Metropolis algorithm. The state of the system is now a vector $\vec{s}$, that contains the states $s_1, s_2, \cdots, s_R$ at the inverse temperatures $\beta_1, \beta_2, \cdots, \beta_R$. As the replicas are not interacting directly, the joint probability

distribution can be factorized into the probability distribution of the individual replicas:

$$P(\vec{s}) = \prod_{i=1}^{R} P_i(s_i) \tag{2.22}$$

where $P_i(s_i)$ is the probability for the replica i with β_i to be in the state s_i. The detailed balance equation then takes the following form:

$$\frac{P(\vec{s})}{P(\vec{s}')} = \frac{\prod_i P_i(s_i' \to s_i)}{\prod_i P_i(s_i \to s_i')} \tag{2.23}$$

The PT algorithm will use two different kinds of Monte Carlo moves: a classic Monte Carlo algorithm to equilibrate the individual replicas and a swap move where the exchange of the states of two neighboring replicas is proposed.

When using a classic MC algorithm to equilibrate an individual replica j, the move only affects s_j, all other components of $\vec{s}$ stay the same. Therefore all but one term in the detailed balance equation cancel out:

$$\begin{aligned} \frac{P(\vec{s}')}{P(\vec{s})} &= \frac{P_j(s_j \to s_j')}{P_j(s_j' \to s_j)} \\ \Leftrightarrow \frac{P_j(s_j')}{P_j(s_j)} &= \frac{P_j(s_j \to s_j')}{P_j(s_j' \to s_j)} \end{aligned} \tag{2.24}$$

This equation is the normal detailed balance condition (2.6) for the replica j. This means that we can use any MC algorithm that works for the system at a fixed temperature in order to equilibrate the individual replicas.

The situation is a bit more complicated when proposing to swap the states of two neighbouring replicas. Let us assume that before the swap, replica j is in state ν and

replica $j+1$ in the state μ, so that $s_j = \nu\,; s_{j+1} = \mu\,; s'_j = \mu\,; s'_{j+1} = \nu$. Most terms in the detailed balance equation will again cancel out:

$$\frac{P(\vec{s})}{P(\vec{s}')} = \frac{P_j(\nu \to \mu) P_{j+1}(\mu \to \nu)}{P_j(\mu \to \nu) P_{j+1}(\nu \to \mu)} \tag{2.25}$$

$$\frac{P_j(\mu) P_{j+1}(\nu)}{P_j(\nu) P_{j+1}(\mu)} = \frac{P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu)}{P_{j,j+1}(\nu, \mu \longrightarrow \mu, \nu)} \tag{2.26}$$

Where $P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu)$ is a simplified notation for the probability that replica j goes from state μ to state ν and replica $j+1$ from ν to μ. If we assume to be in thermodynamic equilibrium, we can write out the probabilities to find a replica in a certain state and obtain:

$$\frac{P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu)}{P_{j,j+1}(\nu, \mu \longrightarrow \mu, \nu)} = \frac{\exp\left[-\beta_j E_j(\nu) - \beta_{j+1} E_{j+1}(\mu)\right]}{\exp\left[-\beta_j E_j(\mu) - \beta_{j+1} E_{j+1}(\nu)\right]} = \tag{2.27}$$

Here, $E_j(\mu)$ is the energy, the replica j would have if it was in the state μ. As not all replicas necessarily share the same Hamiltonian, $E_j(\mu)$ is not always equal to $E_{j+1}(\mu)$. If we assume that all replicas share the same Hamiltonian (meaning $\forall i, j, \mu : E_j(\mu) = E_i(\mu)$), the equation (2.27) simplifies to

$$\frac{P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu)}{P_{j,j+1}(\nu, \mu \longrightarrow \mu, \nu)} = \exp\left[(\beta_j - \beta_{j+1})(E(\mu) - E(\nu))\right] \tag{2.28}$$

This means that for the general case as described by(2.27) the following acceptance probability will be used:

$$\begin{aligned} P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu) = \min \Big[1, \exp\big\{&\beta_j (E_j(\mu) - E_j(\nu)) \\ &+ \beta_{j+1}(E_{j+1}(\nu) - E_{j+1}(\mu))\big\}\Big] \end{aligned} \tag{2.29}$$

Which for (2.28) simplifies to:

$$P_{j,j+1}(\mu, \nu \longrightarrow \nu, \mu) = \min\left[1, \exp\left\{(\beta_j - \beta_{j+1})(E(\mu) - E(\nu))\right\}\right] \tag{2.30}$$

It is very easy to prove the ergodicity of the algorithm, if we demand that the Monte Carlo algorithms used for the individual replicas are ergodic. Then every state $\vec{s}$ is reachable from every other state $\vec{s}'$ because all the components s_i can be reached from s_i' with a finite amount of steps, using only the MC algorithms equilibrating the individual replicas.

As for the standard Metropolis algorithm, the fact that the algorithm fulfills detailed balance and ergodicity does not necessarily mean that it performs very well. For this reason, it is important to understand when this algorithm performs well. There are two main factors that influence the performance of the algorithm:

- The production of new states at the high temperature replicas
- The transport of these new states to the replica with the target temperature T_R

In order to maximize the first factor, the highest temperature T_1 should be chosen such that the free energy landscape does not contain significant local minima anymore. A safe choice would be $T_1 = \infty \Rightarrow \beta_1 = 0$, but this is usually not necessary. If the critical point of the system is known, $T_1 = T_c$ is a good choice, although in this case it might be necessary to deal with critical slowing down.

The transport of these new states on the other hand is not so easy to optimize, as there are several factors that have to be considered. On one hand the transport will be sped up if the acceptance probabilities for the exchanges are high. If we look at (2.30) this can be either achieved by reducing $(\beta_j - \beta_{j+1})$ and therefore increasing R or by reducing $(E(\mu) - E(\nu))$. This energy difference is small if the energy distributions of two neighbouring replicas have a strong overlap or at least a lot of weight close to each other. To enforce this usually also means making the difference in temperature between the neighbouring replicas small, so that their statistics and therefore their energy distributions are similar. But choosing R too big will also slow down the algorithm. First, it is necessary to simulate more replicas slowing down the algorithm

by a linear factor in R. Second, as there are more replicas between T_1 and T_R, the time a state that is generated at T_1 needs in order to arrive at T_R will also grow with R. Furthermore, between the proposed exchanges, the individual replicas should be well-equilibrated in their local wells , so that their energy distributions are not biased.

Some of these factors can be somewhat mitigated by clever techniques. One example is choosing a set of temperatures that optimizes the acceptance probabilities for the exchanges [10]. Another optimization would be to fine-tune the number of MC steps that are used for equilibrating the replicas individually for each temperature[2].

2.5 Measured quantities

As established in section 2.3, the exponential autocorrelation time is a good indicator for the efficiency of an algorithm. But for an Ising system with a large enough external field H, we expect to find almost no configuration in the high energy state. This means that the auto-correlation function will be dominated by the effects of equilibration within one well. The exponential autocorrelation time τ_{exp} however depends mainly on the transitions between the high energy and the low energy state. The rarer these transitions are, the noisier the autocorrelation function becomes. Extracting τ_{exp} from the data is therefore increasingly difficult. This makes it more and more challenging to measure τ for an increasing external field H, posing a serious problem if we are interested in how the performance of parallel tempering depends on H. Relying on measurements of τ_{exp} would for the given reason not only put an upper limit on the field H we can study, it would also expose us to potential systematic errors stemming from fits to an increasingly noisy autocorrelation function.

We will therefore rely on another quantity for evaluating the performance of parallel tempering. This quantity will be called the relaxation time τ_{rel} and represents the time needed by the PT algorithm to return to thermal equilibrium after a disturbance. It will be measured in the following way:

1. The probability of the replica at the target temperature T_R to be in the low energy state is measured by running the simulation long enough. This quantity will be called $\gamma(\infty)$.

2. All replicas are put in the high energy state at time t=0. This puts the system out of equilibrium, as now all replicas are in the otherwise suppressed state.

3. It is recorded whether t PT sweeps after step 2, the replica at T_R is in the low energy state or not. This is done up to time $t_{\max}$, then step 2 is repeated.

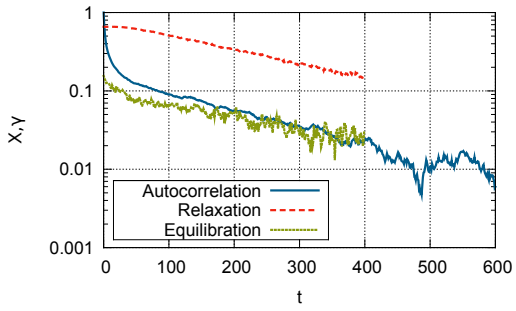
If the external field H is zero, there is strictly speaking no high or low energy state, as both states have exactly the same energy. But the way to measure the relaxation time is exactly the same: we simply call the state with a negative magnetization the high energy state and the one with a positive magnetization the low energy state. We can measure τ_{rel} using the same procedure as for $H > 0$.

By repeating step 2 and 3 often enough, we can calculate the probability $\gamma(t)$ of the replica at T_R to be in the low energy state after t PT sweeps. It is clear that $\gamma(t)$ will approach $\gamma(\infty)$ as the system returns to equilibrium. To analyze the speed of this return to equilibrium, we observe the quantity $\gamma\prime(t) = \gamma(t) - \gamma(\infty)$. It undergoes an exponential decay, similar to the autocorrelation function and by fitting this function with an exponential, it is possible to extract the relaxation time τ_{rel}.

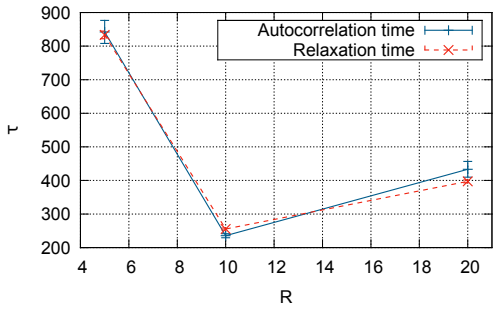
A similar quantity to the relaxation time can be calculated by replacing step 2 with the following:

2' All replicas are put into a random state. This also puts the system out of equilibrium if $H \neq 0$.

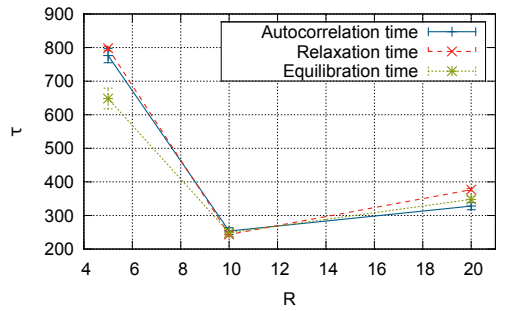
This quantity is called equilibration time and can also be used to study the efficiency of parallel tempering [11]. The equilibration time suffers from a similar problem as the autocorrelation time, as can be seen in figure 2.5a.



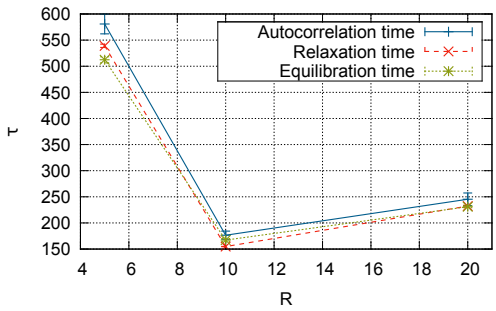
(a) Comparison of the decays for a PT simulation with $R = 10$ and $H = 0.4/L^2$.



(b) Comparison of the different decay times for $H = 0$. (τ_{rel} is omitted because it bears not meaning for $H = 0$)



(c) Comparison of the different decay times for $H = 0.4/L^2$.



(d) Comparison of the different decay times for $H = 1.6/L^2$.

Figure 2.5: Comparison of equilibration time τ_{equ}, autocorrelation time τ_{exp} and relaxation time τ_{rel}. All measurements for an Ising system with $L = 30$ and $T_R = 0.5$.

The relaxation time reveals information about the speed that the PT algorithm is able to reach equilibrium with, if started from a strongly perturbed state. The relaxation time should therefore be related to the exponential autocorrelation time τ_{exp} which is also a measure of the time needed to reach equilibrium. The easier to measure equilibration time τ_{rel} might therefore be a good replacement for the exponential autocorrelation time.

Figure 2.5 shows a comparison between the autocorrelation time, the equilibration time and the relaxation time. The shape of the underlying functions $\chi(t)$ (for the autocorrelation time) and $\gamma'(t)$ (for relaxation and equilibration time) can be seen in figure 2.5a. Although the specific functions all have a different shape, they are subject to a similar exponential decay for $t > 100$. The main difference is that

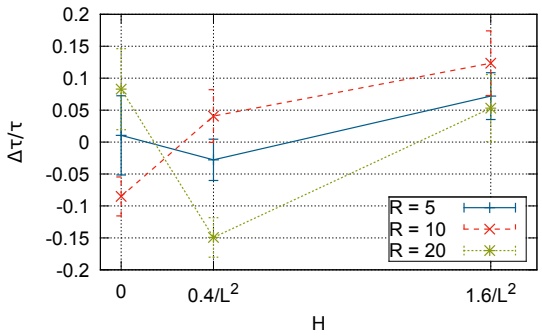


Figure 2.6: Relative difference between τ_{rel} and τ_{exp}

the autocorrelation function as well as the equilibration function will first undergo a rapid decay whereas the relaxation function is constant for small t. As the plot is a logarithmic one, the relative errors become larger as the values of χ or γ drop. When determining the rate of the decay by fitting an exponential function, this means that τ_{rel} will have a much smaller statistical and systematic error.

If we want to use the relaxation time to study parallel tempering, we should make sure that it does not differ significantly from the autocorrelation time. We therefore compare $\tau_{\text{exp}}, \tau_{\text{rel}}$ and τ_{equ} for different values of R and H in the figures 2.5 (b), (c) and (d). It can be seen that the values are quite similar but have statistically significant differences.

The relative difference between the autocorrelation and the relaxation time can be seen in figure 2.6. The number of replicas R seems not to have any clear influence although there are strong fluctuations. Whether there is a strong dependence on the field H can also not be clearly judged. Stronger fields might increase $\Delta\tau/\tau$, but this could also just be a fluctuation, as there is no big increase in the absolute values of $\Delta\tau/\tau$ for larger H. Nevertheless the relaxation time should be taken with a grain of salt when used as an estimate of the autocorrelation time. The relative error seen in figure 2.6 amounts to at about 0.1%, so we should at least add an error margin of 10% when estimating τ_{exp} from τ_{rel}.

Regardless of this, the relaxation time is interesting in its own right, because it is a good portrayal of how the equilibrium distribution is approached when not using the ground state as starting configuration. In more complicated systems, the ground state is not known, therefore all replica will start in an high energy state when compared to the ground state.

CHAPTER 3

PREDICTIONS FOR THE EFFICIENCY OF PARALLEL TEMPERING FOR AN ASYMMETRIC ISING SYSTEM

In order to better understand the results of my measurements, this chapter will present some theoretical considerations on the efficiency of parallel tempering. I was interested in how the efficiency of the algorithm changes when the degeneracy of the ground state is lifted. Therefore this chapter will mainly deal with the efficiency of parallel tempering for either symmetric or strongly asymmetric systems. There is a paper by Machta [11] that deals with exactly this question for a very abstract system and the results and methods of this paper are reviewed in section 3.1. But as the system the paper deals with is highly simplified, it is not immediately clear how its results would transfer to simulations of more complicated systems. In this thesis I will study a 2-dimensional Ising system in order to see whether the results of [11] also apply there.

I will first review the paper by Machta [11] in order to give the reader an understanding of the background of my simulations. I will then explain what quantities I used for studying the 2D Ising system and why I chose them. In section 3.2 I will give theoretical arguments to predict how these quantities might depend on different properties of the parallel tempering algorithm.

3.1 Predictions from a simplified two-well model

In the paper "Strengths and Weaknesses of Parallel Tempering" [11] by Machta the performance of parallel tempering for a simplified model of an double well system

is studied analytically as well as numerically. This chapter will mainly review the predictions made in this paper as they form the backdrop for my own simulations where I try to understand how the predictions generalize to the 2D Ising model.

The system used to make the predictions is an abstracted double well system. Above the critical temperature, there is only one state. Below the critical temperature, there are two states parametrized by σ with $\sigma \in \{0, 1\}$. The free energy is given by:

$$\beta F_\sigma(\beta) = -\frac{1}{2}(\beta - \beta_c)^2(K + H\sigma) \tag{3.1}$$

where K is a parameter controlling the depth of the wells while H controls the difference in free energy between them. If we want the model to represent an Ising system, K would be related to the number of spins and H to the external field. For $H = 0$ the wells are exactly symmetric. For $H > 0$ the symmetry is broken and the well at $\sigma = 1$ will be preferred.

As σ represents a macrostate, the energy at either of the states is not constant but follows a normal distribution with a certain width. The model can therefore be seen as a discrete version of the kind of free energy landscape as seen in figure 2.1a (for $H = 0$) and 2.2 (for $H > 0$). The exact details can be found in [11].

The efficiency of parallel tempering when applied to this model is studied using a mix of analytic and numerical methods. The replicas are placed at evenly spaced inverse temperatures $\beta_1 = \beta_c, \beta_2, \cdots, \beta_R$. For modeling the parallel tempering algorithm, it is assumed that transitions between the both wells are impossible at any temperature below T_c. It is furthermore assumed that every replica is perfectly equilibrated within its well between two proposed exchanges. This also implies that the state σ of the replica at T_c is chosen randomly before every proposed exchange.

It is found that the dynamics governing the efficiency of the algorithm differ significantly between the symmetric ($H = 0$) and the asymmetric ($H > 0$) case. In

section 2.4 two main influences on the performance of parallel tempering were named. The first is the production of new states at high temperature replicas, the second the efficiency of the transport of these newly generated states to the target temperature T_R. In the model, the production of new states only happens at T_c, but there it does happen with perfect efficiency, as a new state is randomly drawn at every step of the algorithm. This means that only the transport of these states will influence the performance of the algorithm.

For $H = 0$, all states have the same free energy and thus there is no preferred direction in the transport of states. The transport is governed by a random walk of the newly generated states to the target temperature T_R. The average time needed for this walk is called the mean first passage time. If an exchange would occur at every time step, it can be computed to be $(R-1)^2$ [17]. This random walk is slowed down by the fact that every exchange is only accepted with the probability P_{acc}. For $H = 0$, this acceptance probability is constant for all pairs of replicas, yielding $\tau(R) \approx \frac{(R-1)^2}{P_{\text{acc}}}$. Analytically evaluating P_{acc} and then optimizing τ as a function of R leads to the following results for $H = 0$:

$$R_{\text{opt}} = 1 + 0.594(\beta_0 - \beta_c)\sqrt{K} \tag{3.2}$$

$$\begin{aligned} \tau &\sim (R_{\text{opt}} - 1)^2 \\ &\sim K(\beta_0 - \beta_c)^2 \end{aligned} \tag{3.3}$$

As these timescales are due to the transport being a diffusive process, this will be called the diffusive autocorrelation time $\tau^d \sim K$.

For $H > 0$, the acceptance probabilities will have a bias towards moving low energy states $\sigma = 1$ to lower temperatures and high energy states $\sigma = 0$ to higher temperatures. This means that the transport is not only a diffusive transport but also has a drift component. The strength of this drift is the velocity with which a low

energy state moves toward lower temperatures. It can be determined by studying the exchange probabilities for the case, where one of the two replicas is in the high energy state, the other one in in the low energy state. We can calculate a mean velocity a low energy state would move towards T_R with, if all other states are occupied by high energy states. This velocity V is calculated by subtracting the acceptance probability for the low energy state moving to a higher temperature from the probability of it going to a lower temperature.

With growing H there is a crossover from diffusion to ballistic motion. For small H, we are in the regime of biased diffusion, but with H large enough it should be possible to describe the motion as a ballistic one with velocity V. Analyzing the exchange probabilities of the model, it can be shown that for H big enough, the number of replicas can be chosen in a way that ensures that $V \approx 1$. The time a low energy state would need for traveling from T_c to T_R would then equal $(R-1)$. It is assumed that $R_{\text{opt}} \sim (\beta_0 - \beta_c)\sqrt{K}$, as for the case with $H = 0$. Therefore the autocorrelation time in the ballistic regime should follow the equation

$$\tau \sim R_{\text{opt}} \sim (\beta_0 - \beta_c)\sqrt{K}. \tag{3.4}$$

3.2 Predictions for the relaxation time

3.2.1 Relaxation times for a non-interacting random walk in one direction

To better understand the results of the simulation, let us first look at the behavior of the relaxation time in a strongly simplified model. This model is supposed to represent the dynamics governing parallel tempering for a strongly asymmetric two-well system. It certainly makes some questionable assumptions, but will nevertheless be able to explain the overall shape of $\gamma'(t)$. We know that quantity $\gamma(t)$ is the probability of the lowest energy replica being in the low energy state, if all replicas have been but into the high energy state t PT-sweeps earlier. As for a strong external

field, $\gamma(\infty) \approx 1$, the quantity $\gamma'(t)$ will be $\gamma'(t) = 1-\gamma(t)$. This is equal the probability to be in the high energy state at time t.

For the simplified model, we are only concerned with exchanges that swap places of a low and a high energy state. Let us make the following assumptions:

- Low energy states can only move to replicas lower in temperature. They never move to a higher temperature. This is surely not realistic for most cases, but we will later replace it by a more realistic condition.

- The acceptance probability for moving a low energy state out of the replica at the highest temperature T_1, to the replica at T_2 is p'.

- For all other replicas, $P_{\text{acc}} = p$ is constant as long as it leads to a low energy state moving down in temperature (and therefore a high energy moving up in temperature).

- Transitions between the two states only occur at the highest temperature replica at T_1. The probability that this replica is in the low energy state is $P_{\text{create}} = \frac{1}{2}$.

All replicas are in the high energy state at the time $t = 0$. Therefore every low energy state that could reach T_R has to be created at T_1 first. Using the parameters described in section 4.1, we assume that every PT sweep consists of equilibrating all individual replicas, then proposing R exchanges. This means that in every sweep we create a low energy state at T_1 with the probability 0.5, followed by the proposed exchanges.

Let us first look at the probability of a newly created low energy state to leave the replica at T_1 during a PT sweep. As there are $R-1$ neighboring pairs of replicas, the probability a specific exchange being proposed is $\frac{1}{R-1}$. Therefore the probability to leave the highest temperature replica can be written the following way:

$$P(\text{leaves}) = 1 - P(\text{stays}) = 1 - \left(1 - \frac{p'}{R-1}\right)^R. \tag{3.5}$$

Any state that does not leave the replica at T_1 during a sweep will be destroyed by the equilibration process at the beginning of the next sweep. The probability that a new low energy state enters the system during a PT sweep is therefore

$$\begin{aligned} P_{\text{new}} =& P_{\text{create}} P(\text{leaves}) \\ =& \frac{1}{2}\left(1 - \left(1 - \frac{p'}{R-1}\right)^R\right). \end{aligned} \tag{3.6}$$

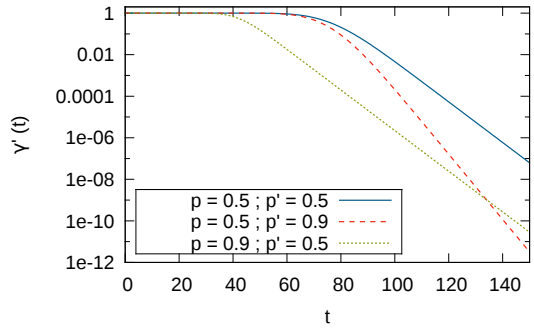
We also need to calculate the probability for a existing low energy state to move a certain number of steps during a given number of PT sweeps. This can be imagined to be similar to a random walk, where the walker can only move in one direction and moves with a probability of $P_{\text{move}} = \frac{p}{R-1}$. The probability for it to have moved l steps after t sweeps is therefore

$$\begin{aligned} P_t(l) &= \text{Binomial}\left(l, t \cdot R, \frac{p}{R-1}\right) \\ &= \binom{tR}{l}\left(\frac{p}{R-1}\right)^l \left(1 - \frac{p}{R-1}\right)^{tR-l}. \end{aligned} \tag{3.7}$$

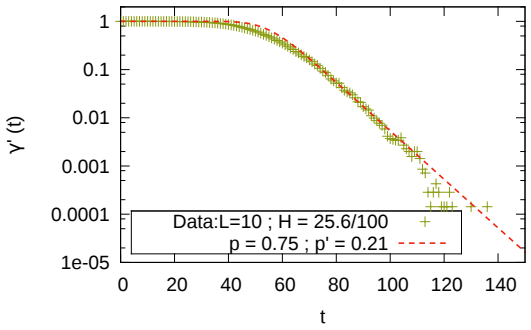
A low energy state that has already left the highest temperature replica has to do $R-2$ steps before it arrives at the lowest temperature. Once it arrives, it will be stuck there. The chance for it to have arrived after t sweeps is therefore chance for the random walk to have moved $R-2$ or more steps.

$$P_{\text{arrived}}(t) = \sum_{l=R-2}^{Rt} \text{Binomial}\left(l, tR, \frac{p}{R-1}\right) \tag{3.8}$$

We can now put all everything together: The probability of the replica at the target temperature to be in the low energy state is the probability of such a state having been created at T_1 and then having moved the steps to the temperature T_R. We assume that these states are created seldom enough to ignore all interactions



(a) $\gamma'(t)$ for $R = 40$ and different parameters p, p'



(b) Comparison of (3.9) with data from PT

Figure 3.1: Numerical evaluation of (3.9) for different parameters and comparison with data

between them[1]. The chance of the replica at T_R to still be in the high energy state at time t is then

$$\begin{aligned}\gamma'(t) &= \prod_{i=1}^{t-1} \big(1 - p_{\text{new}} \cdot p_{\text{arrived}}(t-i)\big). \\ &= \prod_{i=1}^{t-1} \left(1 - \frac{1}{2}\left(1 - \left(1 - \frac{p'}{R-1}\right)^R\right) \sum_{l=R-2}^{R(t-i)} \binom{(t-i)R}{l} \left(\frac{p}{R-1}\right)^l \left(1 - \frac{p}{R-1}\right)^{(t-i)R-l}\right)\end{aligned} \tag{3.9}$$

The index i represents that at every time i a low energy state could have been created and then wandered to T_R in the remaining $t - i$ timesteps. The full expression is evidently a bit unwieldy, so we will restrict ourselves to a numerical evaluation.

I figure 3.1a a plot of (3.9) can be seen for different sets of parameters. The overall shape of $\gamma'(t)$ is independent of these parameters and consists of a plateau followed by a numeric fall-off. The length of the plateau is influenced by R and p, the strength of the fall-off by p'. The plateau exists because every newly created low energy state

[1]We furthermore assume that during the first sweep, the state only moves from the replica at T_1 to the one at T_2. There is also a chance that it already moves further, but including this makes equation (3.9) even more unwieldy and does not have a significant effect on the outcome.

first has to traverse all R states before it can reach the target temperature. This causes a delay of roughly $R \cdot p$. The exponential decay can be explained by the fact that, in the model, every newly created state that leaves T_1 will eventually reach T_R. This means that in the long run, $\gamma'(t)$ will only depend on the probability that no low energy state has been successfully created:

$$\gamma'(t) \approx (1 - p_{\text{new}})^t = e^{t \cdot \log(1 - p_{\text{new}})}. \tag{3.10}$$

This means that $\gamma'(t)$ decays with a typical timescale of

$$\tau_{\text{rel}} = \frac{1}{\log\left(\frac{1}{1-p_{\text{new}}}\right)}. \tag{3.11}$$

By choosing the parameters p and p' accordingly, (3.9) can fit the shape of data from PT simulations quite well (figure 3.1b). But when looking at the choice of parameters p and p' necessary to do so, it becomes clear that the model does not accurately describe the way PT works. In order to describe the observed data, p' and p have to differ dramatically and need to be much smaller than 1. This is at odds with what is observed for the simulations, where the exchange probabilities do not change strongly with temperature and are also usually higher than 60%. The biggest source of error in the model is the assumption that low energy states move only towards lower temperatures. In reality this is not the case. It is possible to partially amend this by giving the variable p the meaning of an average velocity V with which low energy states move. This only influences the plateau observed for low values of t however. As we want to estimate the autocorrelation time, we should instead focus on explaining the exponential falloff of $\gamma'(t)$.

3.2.2 Relaxation time for non-interacting first passage processes

As mentioned before, every low energy state that is created and does not return to T_1 will eventually end up at T_R. Let us therefore reduce the problem to the question whether a successfully created low energy state μ ends up at T_R or returns to T_1. The exchanges are evidently biased towards moving μ to a lower temperature, but as the starting position of μ is next to T_1, its random walk might still lead it back there. The question is therefore whether a discrete space biased random walk will exit a finite interval to the left, ending up at T_1, or the right, ending up at T_R. A formula that describes the probabilities can be found in section 2.4.2.1 of [17]. The probability that a random walker will leave the interval $\{1, 2 \cdots, N-1\}$ to the left is ϵ_-, to leave it to the right ϵ_+.

$$\epsilon_-(x) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^E}{1 - \left(\frac{q}{p}\right)^E} \tag{3.12}$$

$$\epsilon_+(x) = \frac{1 - \left(\frac{q}{p}\right)^x}{1 - \left(\frac{q}{p}\right)^E} = 1 - \epsilon_-(x) \tag{3.13}$$

The parameter x describes the starting position, $q = 1 - p$ the probability to make step to the left. The probability p for a low energy state to make a step to the right is assumed to be constant for all pairs of replicas. If the external field H is zero[2], the random walk is unbiased ($p = 0.5$) and the ϵ_+ is described by the following formula (also [17]):

$$\epsilon_+(x) = \frac{x}{R} \tag{3.14}$$

[2]As already mentioned in 2.5, there is strictly speaking no low energy state for $H = 0$. But if we simply label the state with positive magnetization as "low energy state", we can nevertheless measure the relaxation time the same way we would for $H > 0$.

The probability that a newly created low energy state arrives at T_R is $\epsilon_+(1)$. The probability r that in a given PT sweep a low energy state that eventually reaches T_R is created is therefore for $H \neq 0$[3]:

$$\begin{aligned} r =& p_{\text{new}} \cdot \epsilon_+(1) \\ =& \frac{1}{2} \cdot \left(1 - \left(1 - \frac{p}{R-1}\right)^R\right) \cdot \frac{2 - \frac{1}{p}}{1 - \left(\frac{1}{p} - 1\right)^R}. \end{aligned} \tag{3.15}$$

If $p > 0.5$ and $R \to \infty$, this results in:

$$r \approx \frac{1}{2}\left(1 - e^{-p}\right)\left(2 - \frac{1}{p}\right). \tag{3.16}$$

Using (3.14) for $H = 0$ instead results in

$$r = \frac{1}{2} \cdot \left(1 - \left(1 - \frac{0.5}{R-1}\right)^R\right) \cdot \frac{1}{R} \tag{3.17}$$

$$\approx \frac{(1 - e^{-0.5})}{2R} \qquad [\text{for } R \to \infty] \tag{3.18}$$

The predicted relaxation time is then

$$\tau_{\text{rel}} = \frac{1}{\log\left(\frac{1}{1-r}\right)}. \tag{3.19}$$

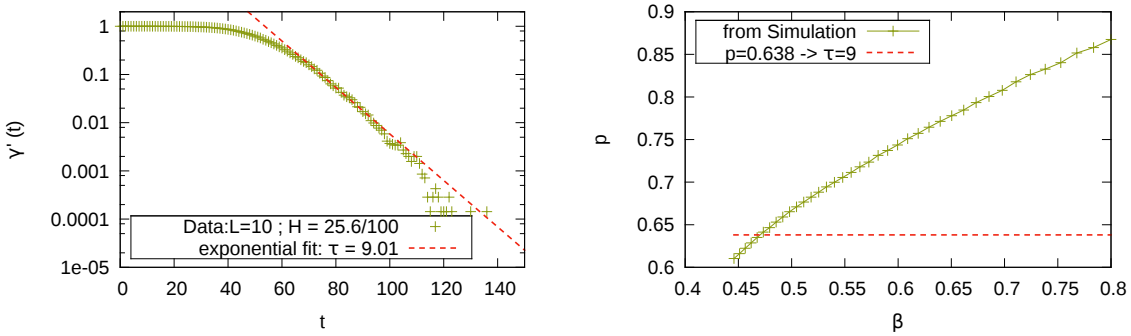
with r from (3.15) or (3.17) respectively. In the case of $H = 0$ we can use some Taylor expansions to better understand the limiting case of $R \to \infty$:

---

[3]In this model, p is the probability to make a step to the left as opposed to making one to the right. It is therefore not P_{acc} from (2.30). For simplicity, (3.15) assumes that both are equal however. Otherwise $\frac{p}{R-1}$ has to be replace by $\frac{p'}{R-1}$ with p' being the acceptance probability for moving from T_1 to T_2. Looking at the data, the approximation seems reasonable for the systems analyzed in this thesis.

$$\tau_{\text{rel}}^{H=0} \approx \frac{1}{\log\left(\frac{1}{1-\frac{c}{R}}\right)} \approx \frac{1}{\log\left(1+\frac{c}{R}\right)} \approx \frac{R}{c} \tag{3.20}$$

where c is $\frac{1-\exp(-0.5)}{2}$.



(a) Exponential fit to measure τ_{rel}

(b) p measured in simulation

Figure 3.2: Comparison of the prediction with data from a PT simulation with a system length of $L = 10$ and $R = 40$ replicas. From $\tau_{\text{rel}} = 9$ we can estimate $p = 0.638$ using (3.15) and (3.19). In (b) this is then compared to $p(\beta)$ measured during the simulation.

In figure 3.2 we compare the predictions from (3.15) to data from a PT simulation. In 3.2b it can be seen that due to p varying with T, an exact prediction for τ_{rel} can not be made. It is however possible to determine which value of p would predict the measured value of $\tau = 9$. Through (3.19) and (3.15) we compute $p = 0.638$ and can compare this to the data. It lies between the maximal and the minimal observed value for p, but much closer to the value of p measured at T_1. This would make sense in the model, as the probability for a low energy state to escape from its starting point would be largely determined by the exchange rates close to T_1.

As the data used for figure 3.2 is the same as the one used for figure 3.1b, we can also see that the estimated value of $p' = 0.21$ is far from the real value. In figure 3.2b we see that $p \approx 0.61$ at T_1. The value of $p = 0.75$ estimated in figure 3.1b seems to be a good estimate of the mean value of the observed p. It should be noted however

that predicting τ_{rel} from figure 3.2b would be very hard. The prediction for τ_{rel} would vary between $\tau = 11$ for $p = 0.61$ (minimal observed p) and $\tau = 3.5$ for $p = 0.87$ (maximal observed p).

3.2.3 Upper bounds on the relaxation time

It should be noted that this prediction is based on three assumptions that will be violated in the PT simulations:

- The value of p is constant for all pairs of replicas. A plot of how p changes with temperature of the replicas can be seen in figure 3.2b.
- A state that has reached T_R will get stuck there. Especially for $\frac{1}{2} \leq p \ll 1$ a low energy state that has reached T_R could migrate back to lower temperatures and might even get destroyed at T_1 again.[4]
- There are no interactions between low energy states. In the simulations this will be violated because there will often be more than two low energy states in the system at the same time.

Let us focus on the violation of the second assumption and how it affects the predictions: Although two low energy states can exchange their place during a PT sweep, this exchange will not have any overall effect because both states are indistinguishable[5]. The system looks the same before and after the exchange of two low energy states and this exchange might as well not have happened. They can therefore be seen as reflective boundaries for each other.

As every low energy state acts as a reflective boundary for other low energy states, (3.19) will underestimate the value of τ and should therefore be seen as a lower bound

[4]Numerical studies suggest that without the "stickiness" condition $\tau_{\text{rel}}(R) \propto R^2$ for $H = 0$.

[5]Two low energy states are only indistinguishable if equilibration within each well is guaranteed between PT sweeps

instead of a strict prediction. An upper bound can be gained by again looking at a first passage process, but this time considering the average time a walker will spend in the system before hitting one of the boundaries. This time is $\mathcal{O}(R^2)$ for $H = 0$ and $\mathcal{O}(R)$ for $H > 0$. But this again ignores the interaction between multiple low energy states in the system. It would be an accurate estimate of τ if we allowed only one low energy state μ in the system at the same time[6]. But the other low energy states in the system can not only block μ from moving to lower temperatures, therefore increasing the time needed to reach T_R. They can also act as a barrier in the other direction, preventing μ from backtracking to higher temperatures, therefore lowering the time needed to reach T_R. These times should thus be seen as an upper limit.

3.2.4 Conclusions

In Conclusion, we predict $\gamma'(t)$ to have the shape of a plateau followed by an exponential decay. The strength of this decay is the relaxation time. If we assume p to remain constant, limits for the relaxation time are the following:

- For $H = 0$ the relaxation time τ_{rel} is theoretically bounded the following way: $\mathcal{O}(R) < \tau_{\text{rel}}(R) < \mathcal{O}(R^2)$. Numerical studies even suggest $\tau_{\text{rel}}(R) = \mathcal{O}(R^2)$.
- For $H > 0$ the relaxation time is bounded by: $\mathcal{O}(1) < \tau_{\text{rel}}(R) < \mathcal{O}(R)$

These limits are based on strongly simplified view of parallel tempering, but they will provide a backdrop for judging the results of the simulations. Any strong deviation from the lower limit will most likely be the result of interactions between the created low energy states. For systems where the production of new low energy states is strongly suppressed, τ_{rel} should be closer to the lower limit. Another way to lower τ_{rel} towards the lower limit might be having more than one replica at each temperature. This would reduce the effect of low energy states acting as barriers for each other.

[6]This could for example be enforced by forbidding the creation of new low energy states at T_1 as long as there is at least one low energy state in the system.

CHAPTER 4

MEASUREMENTS AND RESULTS

4.1 Measurement

The simulated system is a modified two-dimensional Ising model. The Hamiltonian of the system is the following:

$$E = -J \sum_{\langle i,k \rangle} s_i s_k - h' \left(\frac{1}{\beta} - \frac{1}{\beta_c} \right) \sum_i s_i \tag{4.1}$$

When compared to the usual Hamiltonian of an Ising system, the only difference is the temperature dependence of the external field. The way this modification works is that at $\beta = \beta_c$ the external field vanishes and the model becomes completely symmetric. It makes sense to normalize h' such that at the target temperature T_R, the field strength $h' \left(\frac{1}{\beta_R} - \frac{1}{\beta_c} \right)$ equals a the value h.

When simulating this system using a parallel tempering algorithm, each replica has a different Hamiltonian. It is therefore necessary to use equation (2.29) for the acceptance probabilities. With the upper index denoting the replica and the lower index denoting the state for which the value is calculated, the following acceptance probabilities result for (4.1):

$$\begin{aligned} P_{j,j+1}(\mu, \nu \to \nu, \mu) &= \min \left[1, \exp \left\{ \beta^l \left(-JB_\nu - h^l M_\nu + JB_\mu + h^l M_\mu \right) \right. \right. \\ &\qquad \left. \left. + \beta^h \left(-JB_\mu - h^h M_\mu + JB_\nu + h^h M_\nu \right) \right\} \right] \\ &= \min \left[1, \exp \left\{ J \Delta\beta \Delta B + \Delta M \left(h^l \beta^l - h^h \beta^h \right) \right\} \right]. \end{aligned} \tag{4.2}$$

$M = \sum_i s_i$ is the magnetization and $B = \sum_{\langle i,j \rangle} s_i \cdot s_j$ is the number of satisfied bonds minus the number of unsatisfied ones.

The temperatures of the replicas are evenly spaced between the critical temperature[1] $T_c = T_1$ and the target temperature T_R. Each PT sweep consists of 5 sweeps for each of the individual replicas, followed by R proposed exchanges between randomly chosen pairs of neighbouring replicas.

All the individual replicas are simulated using helical boundary conditions (described in section 13.1.1 of [14]). In order to reduce the effect of critical slowing down, the replica at $\beta_1 = \beta_c$ is simulated using a Wolff algorithm [21] whereas all other replicas use the standard Metropolis Monte Carlo algorithm as described in section 2.2.2.

4.2 Results

The simulations were run for various sets of the following parameters:

- The lattice length L. As the lattice is two-dimensional, there are L^2 spins and $2L^2$ bonds in a lattice with length L.
- The number of replicas R used for the parallel tempering algorithm.
- The external field h at T_R. In order to be comparable with the simplified two-well model described in section 3.1, we introduce the quantity $H = hL^2$. This quantity H is equal to the energy difference between a completely unmagnetized ($M = 0$) and a completely magnetized ($M = L^2$) system.

We distinguish between the case of no field ($H = 0$) and the case of a strong field ($H \gg \beta$) where the energy difference between the two states is much bigger than the temperature. As example of the latter, we will usually use $H = 12.8$.

[1] As there is no field at T_c, the critical temperature is the one of the ferromagnetic 2D Ising model without field: $T_c = \frac{2}{\log(1+\sqrt{2})}$.

The results are compared to the predictions from the simplified two-well model reviewed in section 3.1 and to the predictions made in section 3.2. All deviations from the predictions derived from these simplified models should have one of the following reasons:

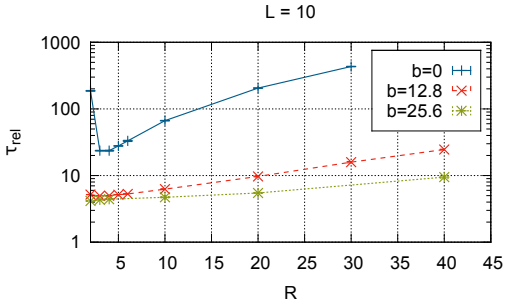
- All predictions assume that the only information conserved in between two PT sweeps are the macrostates of the replicas. But as we only spend a limited amount of time equilibrating the replicas between the PT sweeps, the equilibration is not perfect. This is especially true close to the critical temperature, where critical slowing down drives up the equilibration time. In general, equilibration will be slower for bigger system sizes L^2. The number of sweeps within each replica is fixed to 5 however. For large values of L we would therefore anticipate higher equilibration times than predicted in chapter 3.

- For the predictions, we assumed the acceptance probabilities to be uniform for all neighbouring pairs of replicas. In the simulations, acceptance probabilities were usually much lower close to the critical temperature and became higher at lower temperatures. This comes as no surprise because $\frac{dE}{d\beta}$ diverges at the critical point[2]. The equilibrium ensemble therefore change rapidly as function of the temperature, leading to only small overlaps in energy for neighbouring replicas[3]. Additionally, critical slowing down hinders the equilibration of the individual replicas close to T_C. This can be seen clearly in figure 4.6.

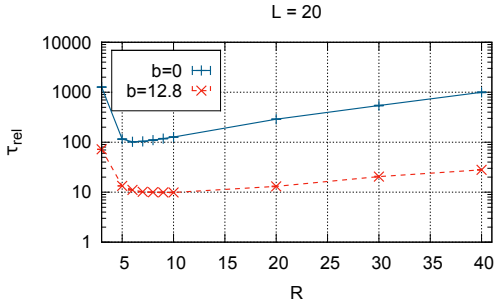
---

[2]Technically, $\frac{dE}{d\beta}$ can of course only diverge for infinite system sizes. But even for finite systems, it has a sharp peak at β_c.

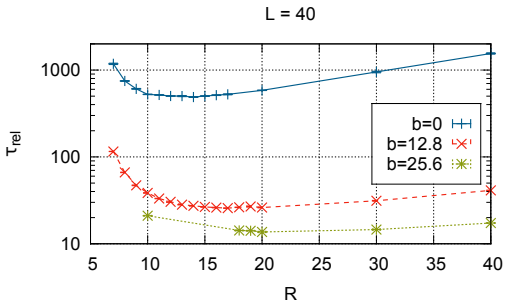
[3]This point is a bit more subtle than it appears. Whereas it is clear that the average energy difference between neighboring replicas does diverge because $\Delta E \approx \Delta\beta \frac{dE}{d\beta}$, this would not necessarily lead to small overlaps, because the variance of the energy also diverges: $\mathrm{Var}(E) = -\frac{dE}{d\beta}$. The width of the energy distribution only diverges with $\sqrt{\mathrm{Var}(E)}$ however, so the distance between the energy distributions of neighboring replicas diverges faster than their width. This means that their overlap tends to zero close to the critical point.



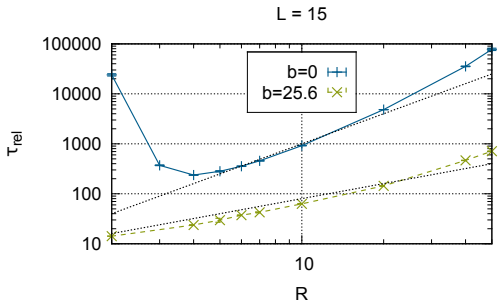
(a) Log-linear plot for $L = 10$.



(b) Log-linear plot for $L = 20$



(c) Log-linear plot for $L = 40$. The minimum is much wider than for $L = 10$ or $L = 20$.



(d) Log-Log plot of τ_{rel} for $L = 15$. The two black dotted lines are two power law curves, R^1 and R^2.

Figure 4.1: The relaxation time τ_{rel} against the number of replicas.

- All predictions were made under the assumption, that transitions between the two states could only be made at the critical temperature T_c. In simulations, transitions are also possible at other temperatures close enough to T_c and the relaxation time could be slightly lowered by this. However as critical slowing down affects the replicas close to T_c it is very unlikely that these transitions occur frequently enough to have a significant impact on τ_{rel}. Especially for larger systems, only the replica at T_c should be effectively able to cross the barrier between the two states[4].

4.2.1 Influence of the number of replicas on the Relaxation Time

The first property we will look at is the relaxation time τ_{rel} for a system of a certain size. Figure 4.1 shows the plots of τ_{rel} against the number of replicas R. It is immediately visible that the presence of an external field shortens the equilibration times considerably. Furthermore it is visible that the minimum of $\tau_{\text{rel}}(R)$ becomes much broader for larger system sizes L^2.

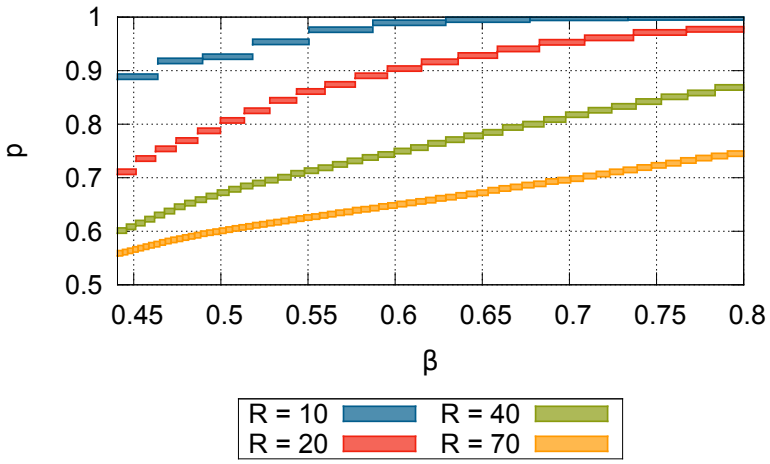


Figure 4.2: The probability p for a low energy state to move towards a lower temperature depends on the replica. It is in general lower for replicas closer to the critical temperature. All curves in this plot come from simulations for $L = 15$ and $H = 25.8$.

In section 3.2, predictions were made for the performance of the PT algorithm and it makes sense to compare the data to these predictions. In figure 4.1d we can see two dotted lines that allow comparing the data to power laws. The lower line is a power law with exponent 1, the upper one with exponent 2. According to the predictions, τ_{rel} should at most grow linear in R for $H > 0$ and at most quadratic for $H = 0$. The curves for the data can however be seen to clearly rise faster than these power laws. At first it might seem strange that the theoretical bounds are violated, but it should not come as a surprise. The predictions in section 3.2 do rely on a constant value of

[4]It should be kept in mind that the replica at T_c is simulated using the Wolff algorithm which does not suffer from critical slowing down.

the probability p for a low energy state to move towards a lower instead of a higher temperature. But if we increase the number of replicas R and keep the rest of the system unchanged, this probability p will significantly decrease. A plot of p over the inverse temperature β can be seen for different values of R in figure 4.2. Especially for β close to the critical point $\beta_c \approx 0.44$, p decreases with increasing R. As the exchange rates close to β_c are estimated to have the biggest influence on τ_{rel}, we can expect τ_{rel} to rise much faster than linear in R. It should be noted that the linear increase in p visible in figure 4.2 for $R = 70$ most likely stems from the fact that the replicas are evenly spaced in temperature. Their distance in β therefore grows for β closer to $T_R = 0.8$. Increasing $\Delta\beta$ will in general lead to a stronger preference for moving low energy states to lower temperatures. In (4.2) we can see that the influence of ΔM grows for larger differences in h and β.

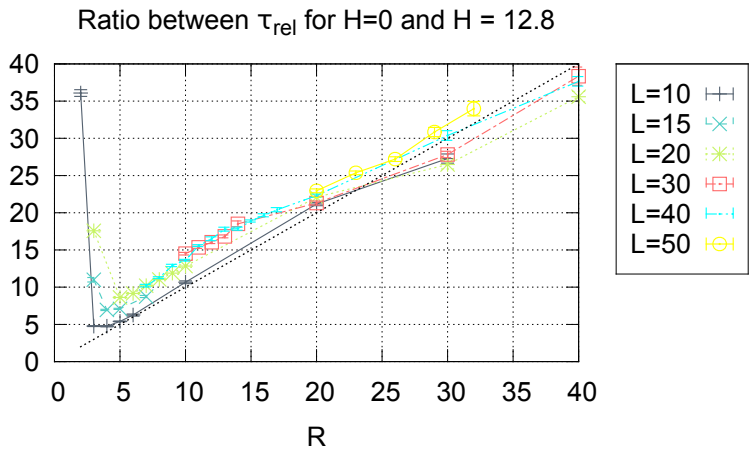


Figure 4.3: The speedup factor of τ_{rel} for a strong field H increases with R

A very interesting value is the factor by which τ_{rel} decreases in the presence of a strong field. In figure 4.3 this speedup factor is plotted against the number of replicas R and compared to a linear curve. The linear curve seems to describe the data quite well for all system sizes. This factor is in agreement with the predictions, even though

the data for τ_{rel} itself is not. The effects that increase the relaxation time therefore seem to affect the system regardless of the presence of a strong field.

4.2.2 The optimal number of replicas as a function of the system size

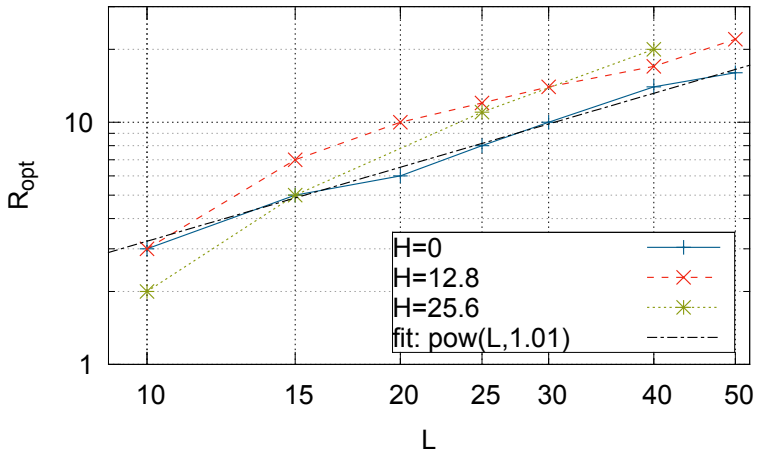


Figure 4.4: Log-log plot of the value of R that optimizes τ_{rel}. The fit is done for $H = 0$ and $L > 20$.

Let us now look at the number of replicas that optimizes the algorithm for a fixed value of L and H. A plot of the choice of R that minimizes the relaxation time can be seen in 4.4. In the absence of an external field ($H = 0$) the curve seems to be linear in R. As R_{opt} is necessarily an integer, the data has a very low granularity. Nevertheless fitting it with a power law works very well (reduced χ^2 is 0.3), retrieving the following dependence:

$$R_{\text{opt}}(L, H = 0) = (0.31 \pm 0.07)L^{1.01\pm0.06}. \tag{4.3}$$

This perfectly fits the prediction from the simplified two-well model (see (3.2)):

$$R_{\text{opt}} \sim \sqrt{K} \sim L^{1.00}. \tag{4.4}$$

Whether a strong external field leaves this exponent untouched and only increases R_{opt} by about 30% is not entirely clear. In figure 4.4 the exponent seems to be unchanged and there also seems to be no clear trend for increasing the field strength. This slightly contradicts the assumptions for the simplified two-well model, where R_{opt} is assumed to be independent of H.

4.2.3 The optimal relaxation time as a function of the system size

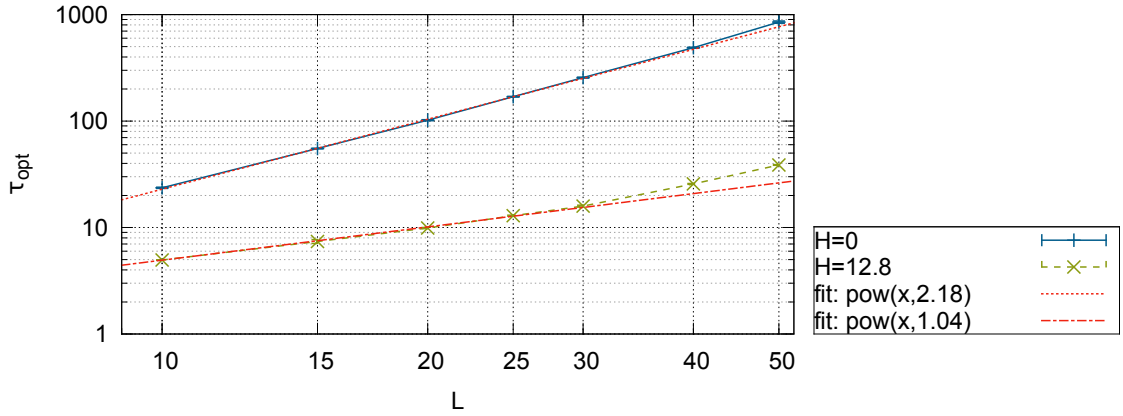


Figure 4.5: Plot of the optimal equilibration time τ_{opt}.

The most interesting question is, how good the parallel tempering algorithm can perform for an optimal choice of R. We are therefore interested in

$$\tau_{\text{opt}} = \min_{R} \left[\tau_{\text{rel}}(R)\right] \tag{4.5}$$

and how it changes with the system size. As to be expected, the results vary between the case with $H = 0$ and with $H \gg 0$. For $H \gg 0$ the result can be described by an almost linear relation between R and τ_{opt}:

$$\tau_{\text{opt}}\big(L, H = 12.8\big) = (0.45 \pm 0.02) L^{1.04 \pm 0.02} \tag{4.6}$$

The data for $H = 0$ agrees with a power law, but in this case with an exponent close to 2:

$$\tau_{\text{opt}}\big(L, H = 0\big) = (0.15 \pm 0.01)L^{2.18\pm0.02} \tag{4.7}$$

In both cases the fit was only done for $L < 30$ and the quality of the fit was not very good (reduced χ^2 of 3 and 4). For $H = 12.8$ the data is in close agreement with the predictions from the simplified two well model, where

$$\tau \sim \sqrt{K} \sim \sqrt{L^2} = L. \tag{4.8}$$

For $H = 0$ the prediction $\tau \sim L^2$ from the simplified two-well model does however not match the data exactly.

Especially for $H \gg 0$, the results for bigger system sizes deviate considerably from the power law. The most probable reason for this is the critical slowing down, that occurs close to β_c. We can investigate this by looking at the exchange probabilities in the PT simulation. Figure 4.6 examines the exchange rates for the systems that were used to generate some of the data points for the $H = 12.8$ curve of figure 4.5. It shows a plot of the probability p that a low energy state moves to a lower temperature, plotted against the inverse temperature β. One interesting feature is that for $L = 50$ and $R = 22$, the value of p actually drops below 50%, whereas for all other curves p steadily increases for higher β. The only reasonable explanation for this seems to be critical slowing down. The replica at β_c uses the Wolff algorithm and is therefore not affected by critical slowing down. But the autocorrelation times for the replicas close to β_c will increase strongly. This leads to a situation where low energy states have problems traversing the replicas close to β_c because they are not properly equilibrated in their replica. There is therefore even a preference for returning back to the critical temperature.

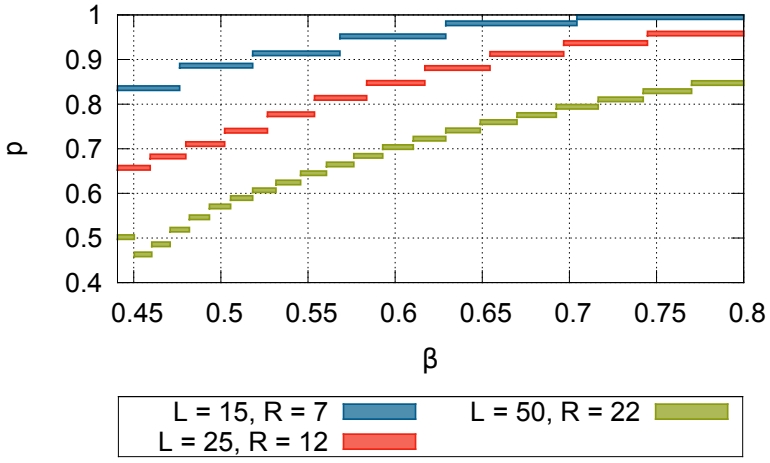


Figure 4.6: Plot of probability p for a low energy state to move to a lower temperature for $H = 12.8$. R is chosen to optimize τ_{rel} for the system size L.

CHAPTER 5

CONCLUSION

Although the findings do in general agree quite well with the results for the simplified two-well model reviewed in section 3.1, there are some deviations for larger system sizes. These deviations do however most probably stem from effects like critical slowing down and uneven exchange rates, not from the increased depth of the free energy well. For simulating systems on a larger scale, one might therefore be well-advised to optimize the efficiency of the algorithm. This can be done by tuning the temperatures of the individual replicas and by optimizing the computational effort used on them in between the PT sweeps. Techniques for this are for example described in [10] and [2].

Regardless of these difficulties for large systems, what one should take home from this is that parallel tempering has a vastly different efficiency depending on the degeneracy of the ground state. When comparing the case where two minima have the same free energy to the case where they do differ significantly in free energy, the algorithm works much better for the latter. As seen in section 4.2.1 relaxation times are decreased by a factor that is roughly equal to the number of replicas. Especially when simulating large systems and using many replicas, one should be aware of this effect to avoid severely underestimating the equilibration times. This is for example highly relevant for spin glasses, where no knowledge about the degeneracy of the ground state is available previous to the simulation.

One subject not touched upon is that the analysis in this thesis assumes that there is no direct cost for increasing the number of replicas in a simulation. But usually, in-

creasing the number of replicas will increase the computational effort for each sweep. Simulating R replicas needs R times as much processing power as would simulating a system at a single temperature. This increase is usually worthwhile as parallel tempering is often by many orders of magnitudes faster than single-temperature Monte Carlo. But it certainly has an influence if we would want to find the number R_{opt} that offers the lowest relaxation time per used processor time.

On the other hand, the parallel tempering algorithm is very easy to parallelize. Because the replicas do only interact sporadically, the algorithm lends itself to be run on different processor cores or even different machines, where each core runs one replica. Having said that, increasing R will actually slow the simulation down from a certain point on. The value R_{opt} discussed in section 4.2.2 therefore provides an upper bound to the degree of concurrency that is possible. A similar algorithm that has no upper bound on concurrency would be for example Population Annealing [6, 12].

BIBLIOGRAPHY

[1] Baxter, R. J. *Exactly solved models in statistical mechanics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1982.

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

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

[4] Brush, S. G. History of the lenz-ising model. *Rev. Mod. Phys. 39*, 4 (1967), 883–893.

[5] Geyer, C. J. Markov chain monte carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (1991), Interface Foundation, pp. 156–163.

[6] Hukushima, K., and Iba, Y. Population annealing and its application to a spin glass. In *The Monte Carlo Method in the Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm* (2003), James E. Gubernatis, Ed., vol. 690, AIP, pp. 200–206.

[7] Hukushima, K., and Nemoto, K. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan 65*, 6 (1996), 1604–1608.

[8] Ising, E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei 31* (1925), 253–258. 10.1007/BF02980577.

[9] Kar, P., Nadler, W., and Hansmann, U. H. E. Microcanonical replica exchange molecular dynamics simulation of proteins. *Phys. Rev. E 80*, 5 (2009), 056703.

[10] Katzgraber, H. G., Trebst, S., Huse, D. A., and Troyer, M. Feedback-optimized parallel tempering monte carlo. J. Stat. Mech. P03018 (2006), 2006.

[11] Machta, J. Strengths and weaknesses of parallel tempering. *Phys. Rev. E 80*, 5 (2009), 056706.

[12] Machta, J. Population annealing: An effective monte carlo method for rough free energy landscapes, 2010.

[13] Metropolis, N., Rosenbluth, M., Teller, A., and Teller, E. Equation of state calculations by fast computing machines. *J.Chem.Phys. 21* (1953), 1087.

[14] Newman, M. E. J., and Barkema, G. T. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, 1999.

[15] Onsager, L. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev. 65*, 3-4 (1944), 117–149.

[16] Parisi, G. Some considerations of finite-dimensional spin glasses. *J. Phys. A: Math. Gen. 41*, 32 (2008), 324002.

[17] Redner, S. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.

[18] Romá, F., Risau-Gusman, S., Ramirez-Pastor, A., Nieto, F., and Vogel, E. 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 (2009), 2821–2838.

[19] Swendsen, R. H., and Wang, J.-S. Replica monte carlo simulation of spin-glasses. *Phys. Rev. Lett. 57*, 21 (1986), 2607–2609.

[20] Swendsen, R. H., and Wang, J.-S. Nonuniversal critical dynamics in monte carlo simulations. *Phys. Rev. Lett. 58*, 2 (1987), 86–88.

[21] Wolff, U. Collective monte carlo updating for spin systems. *Phys. Rev. Lett. 62*, 4 (1989), 361–364.

[22] Yoshioka, D. *Statistical Physics*. Springer Berlin Heidelberg, 2007.