Sign and Complex Action Problems: The Potts Model Approximation to Dense QCD as a Solvable Example[1]

Uwe-Jens Wiese
Institute for Theoretical Physics, Bern University
Sidlerstrasse 5, CH-3012 Bern, Switzerland

May 10, 2006

[1]Lectures given at the Doctoral Training Program, ECT*, Trento, May 8-12, 2006

Abstract

These lectures provide an introduction to the sign problem which prevents the numerical simulation of many important problems in both condensed matter and particle physics. In order to be self-contained, these notes begin with a short introduction to the Ising model and its numerical simulation using the Monte Carlo method. First, the Ising model is introduced and important physical observables are discussed. The phase structure of the Ising model in a sufficiently large number of spatial dimensions is investigated using mean field theory. Then the model is solved analytically in one dimension, and the 2-dimensional model is studied using a duality transformation. A cluster representation of the Ising model is also discussed. In the next step the Monte Carlo method is introduced and the properties of ergodicity and detailed balance are discussed. It is shown that the Metropolis algorithm is ergodic and obeys detailed balance. The much more efficient Swendsen-Wang and Wolff cluster algorithms are also introduced.

Monte Carlo simulations of lattice QCD at non-zero baryon chemical potential μ suffer from the notorious complex action problem. We consider QCD with static quarks coupled to a large chemical potential. This leaves us with an $SU(3)$ Yang-Mills theory with a complex action containing the Polyakov loop. Close to the deconfinement phase transition the qualitative features of this theory, in particular its $\mathbf{Z}(3)$ symmetry properties, are captured by the 3-d 3-state Potts model. We solve the complex action problem in the Potts model by using a cluster algorithm. The improved estimator for the μ-dependent part of the Boltzmann factor is real and positive and is used for importance sampling.

We also examine staggered lattice fermions in $(3+1)$ dimensions with a four-fermion interaction and $\mathbf{Z}(2)$ chiral symmetry. This model suffers from a severe sign problem and cannot be studied with standard fermion algorithms. We a meron-cluster algorithm which solves the sign problem and leads to high-precision numerical data.

Contents

Chapter 1

Motivation

Problems in science or engineering are often rather complex and may not allow for an analytic solution with paper and pencil. In particular, in physics there are numerous important problems that cannot be solved analytically. The corresponding physical systems often contain a large number of strongly coupled degrees of freedom. Examples from classical physics include fluid dynamics (with important practical applications such as weather forecast or climate modeling), but also models from statistical mechanics in general. Quantum physics also gives rise to important complex problems. They range from quantum statistical mechanics to quantum field theory and include very hard problems like understanding high-temperature superconductors (systems of strongly correlated electrons) or solving Quantum Chromodynamics (QCD) — the theory of strongly interacting quarks and gluons.

A very interesting class of physics problems is described by the Ising model — a simple model of classical statistical mechanics. This model describes a wide variety of strongly coupled physical systems — ranging from idealized magnets to certain aspects of the quark-gluon plasma. In one spatial dimension the Ising model can be solved easily using analytic methods. Even in two dimensions certain aspects of the model can be understood analytically. In the physically most relevant 3-d case, however, analytic methods are of limited use. Numerical methods, on the other hand, provide a very powerful tool to solve the Ising model. The Ising model is ideally suited to familiarize ourselves with the Monte Carlo method of numerical simulation. If we are in command of this method, we can

not only solve the Ising model, but a whole variety of models from both classical and quantum statistical mechanics, as well as from quantum field theory.

Non-perturbative dense QCD can presently not be studied from first principles because Monte Carlo simulations of lattice QCD with non-zero baryon chemical potential μ suffer from a severe complex action problem. The Boltzmann factor in the path integral can then not be interpreted as a probability and standard importance sampling methods fail. In particular, when the μ-dependent part of the Boltzmann factor is included in the measured observables, due to severe cancellations the required statistics is exponentially large in the space-time volume.

The complex action problem prevents the numerical simulation of a large class of interesting physical systems including other field theories at non-zero chemical potential or non-zero θ-vacuum angle as well as some fermionic field theories with an odd number of flavors. Another problem in the same category is the so-called fermion sign problem which arises for fermionic path integrals formulated in a Fock state basis. The problem is due to paths that correspond to an odd permutation of fermion positions which contribute negatively to the path integral. There are numerous condensed matter systems ranging from the repulsive Hubbard model away from half-filling to antiferromagnetic quantum spin systems in an external magnetic field that cannot be simulated with standard Monte Carlo algorithms. Meron-cluster algorithms have been used to solve the sign or complex action problems in several of these cases. For example, the first meron-cluster algorithm has led to a solution of the complex action problem in the 2-d $O(3)$ symmetric field theory at non-zero θ-vacuum angle. In this model, some of the clusters are half-instantons — thus the name meron-clusters. The complex action problem also arises in the 2-d $O(3)$ model at non-zero chemical potential. Recently, the meron-concept has been generalized to fermions. Meron-cluster algorithms have led to a complete solution of the fermion sign problem in a variety of models including non-relativistic spinless fermions, relativistic staggered fermions, and some models in the Hubbard model family. Recently, a meron-cluster algorithm has been used to solve the sign problem that arises for quantum antiferromagnets in an external magnetic field.

In Wilson's formulation of lattice QCD the quarks are represented by Grassmann fields. When the quarks are integrated out, they leave behind a fermion determinant that acts as a non-local effective action for the gluons. At zero chemical potential and for an even number of flavors, the fermion determinant is real and positive and can thus be interpreted as a probability for generating

gluon field configurations. Despite the fact that standard importance sampling techniques apply, the non-local nature of the effective gluon action makes lattice QCD simulations with dynamical fermions still very time consuming. With a non-zero chemical potential for the baryon number, the fermion determinant becomes complex and standard importance sampling techniques fail completely. This is the reason why non-perturbative QCD at non-zero baryon density can presently not be studied from first principles.

It is natural to ask if a meron-cluster algorithm can be used to solve the complex action problem in QCD. When one integrates out the quarks, this is unlikely because the meron-concept probably does not apply to the non-local effective action for the gluons. On the other hand, when one describes the quarks in a Fock state basis, the complex action problem is complicated by a fermion sign problem. Here we address a simpler problem. We consider QCD in the limit of very heavy quarks with a large chemical potential. Since these quarks are not dynamical, the fermion permutation sign does not arise. Still, there remains a complex action problem due to the non-zero chemical potential. In the effective gluon action the heavy quarks manifest themselves as Polyakov loops — i.e. as non-local composite operators constructed from gluon fields. This non-locality seems more manageable than the one that arises for a general fermion determinant. Here we aim at a complete solution of the complex action problem. However, at the level of the effective gluon theory, we still cannot apply a meron-cluster algorithm, because the construction of efficient cluster algorithms for non-Abelian gauge theories seems to be impossible for Wilson's formulation of lattice field theory.

The 3-d $\mathbf{Z}(3)$-symmetric Potts model has often been used as an approximation to QCD with static quarks. In particular, the phase transition to a broken $\mathbf{Z}(3)$ symmetry phase at high temperature corresponds to the first order deconfining phase transition in QCD. A term that corresponds to a chemical potential can also be included in the Potts model, explicitly breaking the $\mathbf{Z}(3)$ symmetry. The first order deconfinement phase transition persists but it becomes weaker and ultimately disappears in a critical endpoint. This point is expected to be in the universality class of the 3-d Ising model. In principle, one can imagine to derive an effective 3-d 3-state Potts model directly from QCD by integrating out all degrees of freedom except for the $\mathbf{Z}(3)$ phase of the Polyakov loop. However, the resulting Potts model action would be very complicated and cannot be derived in practice, except in the strong coupling limit. Here we approximate QCD with heavy quarks by a 3-d $\mathbf{Z}(3)$-symmetric Potts model with a standard

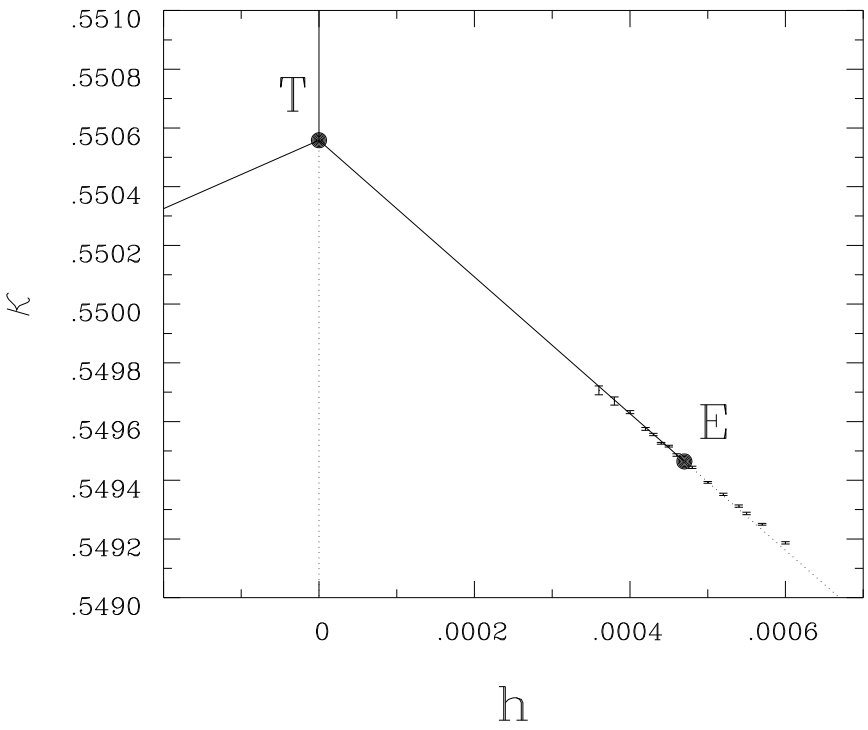

Figure 1.1: *The phase diagram in the (h, κ)-plane. The ordinary deconfinement phase transition at $h = 0$ is a triple point T from which a line of first order phase transitions emerges. This line terminates in the critical endpoint E.*

nearest neighbor interaction. Universal features like the nature of the critical endpoint of the deconfinement phase transition are correctly reproduced in this approximation. Figure 1.1 contains the phase diagram of the 3-d 3-state Potts model in the (h, κ)-plane. The parameter κ is the standard Potts model coupling and the parameter h represents $\exp(-\beta(M - \mu))$ in QCD with heavy quarks of mass M at chemical potential μ and inverse temperature β. The ordinary first order deconfinement phase transition at $h = 0$ (point T in fig. 1) extends into a line of first order transitions that terminates in the critical endpoint E. The numerical data for the Potts model with complex action are consistent with 3-d Ising behavior and allow us to extract critical exponents.

In the absence of a chemical potential, the Potts model can be simulated with the original Swendsen-Wang cluster algorithm. When a chemical potential is introduced, the Potts model still suffers from the complex action problem and

standard importance sampling methods including the cluster algorithm fail. An improved estimator for the μ-dependent part of the Boltzmann factor is constructed by averaging analytically over all configurations related to each other by cluster flips. In contrast to the original Boltzmann factor, the improved estimator is real and positive and can be used for importance sampling. This solves the complex action problem completely.

The numerical simulation of lattice fermions is a notoriously difficult problem which is the major stumbling block in solving QCD and other field theories. The standard method is to integrate out the fermions and to simulate the resulting bosonic problem with a non-local action. In several cases of physical interest — for example, for QCD with an odd number of flavors or with non-zero chemical potential — the bosonic Boltzmann factor may become negative or even complex and can thus not be interpreted as a probability. When the sign or the complex phase of the Boltzmann factor is included in measured observables, the numerical simulation suffers from severe cancellations resulting in a sign problem. The standard fermion algorithms are incapable of exploring such models. As a consequence, QCD is usually simulated with an even number of flavors and at zero chemical potential. Even in the absence of a sign problem, the simulation of fermions is very difficult. For example, lattice QCD simulations suffer from critical slowing down when one approaches the chiral limit of massless quarks. In particular, this makes it difficult to identify the universality class of the finite temperature QCD chiral phase transition.

Even in simple models with four-fermion interactions the identification of the critical behavior is a non-trivial issue. A model with N fermion flavors shows mean-field behavior in the $N = \infty$ limit. On the other hand, at finite N one finds the non-trivial critical behavior that one expects based on standard universality arguments. For example, it has been verified that the chiral phase transition in a (2+1)-d $N = 4$ four-fermion model with $\mathbb{Z}(2)$ chiral symmetry is in the universality class of the 2-d Ising model. Due to the fermion sign problem, standard fermion methods do not work in models with fewer than four flavors.

We will apply a cluster algorithm technique to tackle the sign problem to a model of staggered fermions in $(3+1)$ dimensions. The model has two flavors and a $\mathbb{Z}(2)$ chiral symmetry that is spontaneously broken at low temperatures. Taking the positive square root of the fermion determinant of a four flavor model is not appropriate, because the determinant of the two flavor model can be negative. Because of the sign problem, standard fermion algorithms fail in this case. Our

meron cluster algorithm is the only numerical method available to simulate this model. In this method we do not integrate out the fermions but describe them in a Fock state basis. The resulting bosonic model of fermion occupation numbers interacts locally, but has a non-local fermion permutation sign resulting from the Pauli principle. Standard numerical methods would suffer from severe cancellations of positive and negative contributions to the partition function. The idea of the meron cluster algorithm is to decompose a configuration of fermion occupation numbers into clusters which can be flipped independently. Under cluster flip an occupied site becomes empty and vice versa. Clusters whose flip change the fermion sign are called merons. If a configuration contains a meron-cluster, its contribution to the partition function is canceled when the meron-cluster is flipped. The observables that we consider get non-zero contributions from the zero- and two-meron sectors only. Our algorithm ensures that configurations with more than two merons are never generated, which leads to an exponential gain in statistics and to a complete solution of the sign problem.

Like other cluster algorithms the meron algorithm substantially reduces critical slowing down. This allows us to work directly in the chiral limit. As a result, we can study the nature of the chiral phase transition in great detail. The $\mathbb{Z}(2)$ chiral symmetry is spontaneously broken at low temperatures and gets restored in the high-temperature phase. As expected, the system close to the finite temperature critical point is in the universality class of the 3-d Ising model. We verify this in a high-precision finite-size scaling investigation of the chiral susceptibility.

Chapter 2

The Ising Model

The Ising model is one of the simplest models in classical statistical mechanics and still it is applicable to a wide variety of physical systems. It describes idealized magnets, the mixing of fluids, critical opalescence in boiling water at high pressure, and even special features of the quark-gluon plasma that filled the early Universe.

2.1 Definition and Basic Properties

Let us consider the simplest classical spin model — the so-called Ising model. Here the word spin does not mean that we deal with quantized angular momenta. All we do is work with classical spin variables that take values $s_x = \pm 1$. The Ising spins are located on the sites of a d-dimensional spatial cubic lattice. The Ising model is characterized by its classical Hamilton function (not a quantum Hamilton operator) which simply specifies the energy of any configuration of spins. The Ising Hamilton function includes a sum of nearest neighbor contributions

$$\mathcal{H}[s] = -J \sum_{\langle xy \rangle} s_x s_y - \mu B \sum_x s_x , \tag{2.1.1}$$

with a ferromagnetic coupling constant $J > 0$ that favors parallel spins, plus a coupling to an external magnetic field B. The classical partition function of this

system is given by

$$Z = \sum_{[s]} \exp(-\beta \mathcal{H}[s]) = \prod_x \sum_{s_x = \pm 1} \exp(-\beta \mathcal{H}[s]). \tag{2.1.2}$$

The sum over all spin configurations corresponds to an independent summation over all possible orientations of individual spins. Thermal averages are computed by inserting appropriate observables $\mathcal{O}[s]$. For example, the magnetization is given by

$$\mathcal{M}[s] = \sum_x s_x, \tag{2.1.3}$$

and its thermal expectation value is given by

$$\langle \mathcal{M} \rangle = \frac{1}{Z} \sum_{[s]} \mathcal{M}[s] \exp(-\beta \mathcal{H}[s]) = \frac{\partial \log Z}{\partial(\beta \mu B)}. \tag{2.1.4}$$

The magnetic susceptibility is defined as

$$\chi = \frac{1}{L^d}(\langle \mathcal{M}^2 \rangle - \langle \mathcal{M} \rangle^2) = \frac{1}{L^d} \frac{\partial^2 \log Z}{\partial(\beta \mu B)^2}, \tag{2.1.5}$$

where L^d is the spatial volume. Similarly, the spin correlation function is defined as

$$\langle s_x s_y \rangle = \frac{1}{Z} \sum_{[s]} s_x s_y \exp(-\beta \mathcal{H}[s]). \tag{2.1.6}$$

At large distances the so-called connected spin correlation function typically decays exponentially

$$\langle s_x s_y \rangle_c = \langle s_x s_y \rangle - \langle s_x \rangle \langle s_y \rangle \sim \exp(-|x - y|/\xi), \tag{2.1.7}$$

where ξ is the correlation length. The susceptibility is the connected correlation function summed over all pairs of points x and y, i.e.

$$\chi = \frac{1}{L^d} \sum_{x,y} \langle s_x s_y \rangle_c. \tag{2.1.8}$$

At general temperatures the correlation length is typically just a few lattice spacings. When one models real materials the Ising model would generally be a great over-simplification, because real magnets, for example, not only have nearest-neighbor couplings. Still, the details of the Hamilton function at the scale

of the lattice spacing are not always important. There is a critical temperature T_c at which ξ diverges and universal behavior arises. At this temperature a second order phase transition occurs. Then the details of the model at the scale of the lattice spacing are irrelevant for the long-range physics that takes place at the scale of ξ. In fact, at their critical temperatures real materials behave just like the simple Ising model. This is why the Ising model is so interesting. It is just a very simple member of a large universality class of different models, which all share the same critical behavior. This does not mean that they have the same values of their critical temperatures. However, their correlation lengths diverge at the critical temperature with the same exponent ν, i.e.

$$\xi \propto |T - T_c|^{-\nu}, \tag{2.1.9}$$

their magnetizations go to zero at the critical temperature with the same exponent β

$$\langle \mathcal{M} \rangle \propto |T - T_c|^{\beta}, \; T \leq T_c, \tag{2.1.10}$$

and their susceptibilities diverge with the same exponent γ

$$\chi \propto |T - T_c|^{-\gamma}, \tag{2.1.11}$$

i.e. the critical exponents ν, β, and γ are identical for different systems in the same universality class.

2.2 Mean Field Theory

Understanding critical behavior is a highly nontrivial issue. For example, there is no generally applicable analytic method that allows us to determine critical exponents exactly. An exception are theories in one or two dimensions. As we will see, in 1-d the Ising model can be solved easily. In 2-d a certain symmetry — conformal symmetry — allows exact calculations. In the 3-d case the so-called ϵ-expansion at least provides a systematic expansion whose convergence is, however, not always guaranteed. In the next chapter we will learn about the Monte Carlo method which provides an alternative numerical tool to understand critical phenomena. To illustrate critical behavior we want to use an approximate method — mean field theory. Its results become exact in the limit of infinitely many dimensions, but should not be trusted quantitatively in the physically most relevant case $d = 3$. The idea behind mean field theory is to assume that a spin s_x

that interacts with its fluctuating nearest neighbors s_y can simply be coupled to a constant averaged mean value $\langle s\rangle_m$ for the spin. The mean field approximation to the Hamilton function takes the form

$$\mathcal{H}_m[s] = -J\sum_x s_x d\langle s\rangle_m - \mu B\sum_x s_x = -\mu B_{\rm eff}\sum_x s_x, \tag{2.2.1}$$

where

$$\mu B_{\rm eff} = \mu B + dJ\langle s\rangle_m, \tag{2.2.2}$$

is an effective magnetic field acting on the spin s_x. In the full theory the magnetic field generated by the neighboring spins fluctuates with the values of the spins s_y. In mean field theory one neglects these fluctuations and treats the magnetic field as an averaged constant.

In the mean field approximation the partition function can now be evaluated easily

$$\begin{aligned} Z_m &= \sum_{[s]}\exp(-\beta\mathcal{H}_m[s]) = \prod_x\sum_{s_x=\pm 1}\exp(\beta\mu B_{\rm eff}\sum_x s_x) \\ &= [\sum_{s_{x_0}=\pm 1}\exp(\beta\mu B_{\rm eff}s_{x_0})]^V = [2\cosh(\beta\mu B_{\rm eff})]^V, \end{aligned} \tag{2.2.3}$$

where x_0 is any arbitrarily chosen lattice point and $V = L^d$ is the volume (the total number of lattice sites). The factor d is just the number of bonds per spin in d dimensions. Since $B_{\rm eff}$ depends on the mean field value of the spin $\langle s\rangle_m$ the calculation is not yet complete. The next step is to determine the average spin from a consistency condition. In mean field theory we have

$$\begin{aligned} \langle s\rangle_m &= \frac{1}{Z_m}\sum_{[s]} s_{x_0}\exp(-\beta\mathcal{H}_m[s]) \\ &= \frac{1}{Z_m}[\sum_{s_{x_0}=\pm 1} s_{x_0}\exp(\beta\mu B_{\rm eff}s_{x_0})][2\cosh(\beta\mu B_{\rm eff})]^{V-1} \\ &= \tanh(\beta\mu B_{\rm eff}) = \tanh(\beta(\mu B + dJ\langle s\rangle_m)). \end{aligned} \tag{2.2.4}$$

This is the consistency condition for $\langle s\rangle_m$.

Next we assume that the external magnetic field is switched off, i.e. $B = 0$. Then the consistency condition takes the form

$$\langle s\rangle_m = \tanh(\beta dJ\langle s\rangle_m). \tag{2.2.5}$$

For $d\beta J < 1$ this has only the trivial solution $\langle s\rangle_m = 0$. However, for $d\beta J > 1$ there is, in addition, a solution with $\langle s\rangle_m \neq 0$. This solution describes spontaneous symmetry breaking: without any bias by an external magnetic field, the spins decide collectively to point in a common direction. The nontrivial solution appears at a critical temperature T_c that is given by

$$d\beta_c J = 1 \;\Rightarrow\; T_c = \frac{dJ}{k_B}. \tag{2.2.6}$$

Above this temperature the spin system is in the so-called symmetric (or unbroken) phase with $\langle s\rangle_m = 0$, while for $T < T_c$ the system is in the broken (or ordered) phase in which spontaneous symmetry breaking occurs and $\langle s\rangle_m \neq 0$. The two phases are separated by a phase transition. The value of the average spin determines if the system is in the unbroken or in the ordered phase. Hence, $\langle s\rangle$ is known as an order parameter.

It remains to be shown that the nontrivial solution is indeed physically realized. For that purpose we compare the free energies of the two phases. In the symmetric phase we have $B_{\text{eff}} = 0$ and hence

$$Z_m = [2\cosh(\beta\mu B_{\text{eff}})]^V = 2^V = \exp(-\beta F_s), \tag{2.2.7}$$

In the broken phase, on the other hand, we have $|B_{\text{eff}}| > 0$ and hence

$$\exp(-\beta F_b) > \exp(-\beta F_s) \;\Rightarrow\; F_b < F_s. \tag{2.2.8}$$

Since the free energy of the broken phase is smaller than the one of the symmetric phase, the broken phase is thermodynamically stable for $T < T_c$.

Next we compute the average spin (magnetization per lattice site) for temperatures below but close to T_c. Expanding the consistency condition for small $\langle s\rangle_m$

$$\langle s\rangle_m = \tanh(\beta dJ\langle s\rangle_m) \approx \beta dJ\langle s\rangle_m - \frac{1}{3}(\beta dJ\langle s\rangle_m)^3, \tag{2.2.9}$$

one obtains

$$\beta dJ\langle s\rangle_m = \sqrt{3(\beta dJ - 1)} \;\Rightarrow\; \langle s\rangle_m \propto \sqrt{T_c - T} = |T - T_c|^\beta. \tag{2.2.10}$$

Here β is the critical exponent introduced before. Its mean field value $\beta = 1/2$ is not exact. Since the order parameter goes to zero continuously at T_c, the phase transition is of second order.

At a first order phase transition, on the other hand, the order parameter makes a discontinuous jump. The Ising model undergoes a first order phase transition at $T < T_c$ when on changes the external magnetic field B from positive to negative values. At $B = 0$ the magnetization changes abruptly from $+\langle s\rangle_m$ to $-\langle s\rangle_m$.

2.3 Exact Results for the 1-d Ising Model

Mean field theory nicely illustrates the qualitative behavior of the Ising model but it yields only approximate results. In this section we will derive some exact results for the Ising model in one dimension. The 1-d Ising model is easy to solve analytically. Its partition function (for $B = 0$) is given by

$$Z = \prod_x \sum_{s_x=\pm 1} \exp(\beta J \sum_{\langle xy\rangle} s_x s_y). \tag{2.3.1}$$

We consider a lattice with L sites and with periodic boundary conditions (i.e. $s_{x+L} = s_x$). Introducing bond variables

$$b_{\langle xy\rangle} = s_x s_y = \pm 1, \tag{2.3.2}$$

the partition function can be rewritten as

$$Z = 2 \prod_{\langle xy\rangle} \sum_{b_{\langle xy\rangle}=\pm 1} \exp(\beta J \sum_{\langle xy\rangle} b_{\langle xy\rangle}) \delta_{\prod_{\langle xy\rangle} b_{\langle xy\rangle},1}. \tag{2.3.3}$$

The constraint

$$\prod_{\langle xy\rangle} b_{\langle xy\rangle} = 1, \tag{2.3.4}$$

that is enforced by the Kronecker δ-function is a consequence of periodic boundary conditions. The δ-function can be rewritten as

$$\delta_{\prod_{\langle xy\rangle} b_{\langle xy\rangle},1} = \frac{1}{2} \sum_{m=0,1} (\prod_{\langle xy\rangle} b_{\langle xy\rangle})^m. \tag{2.3.5}$$

Hence, the partition function takes the form

$$Z = \sum_{m=0,1} \prod_{\langle xy\rangle} \sum_{b_{\langle xy\rangle}=\pm 1} \exp(\beta J b_{\langle xy\rangle}) b_{\langle xy\rangle}^m = \{[2\cosh(\beta J)]^L + [2\sinh(\beta J)]^L\}. \tag{2.3.6}$$

This exact results is not identical with the mean field result of eq.(2.2.3).

Let us also compute the correlation function $\langle s_x s_y \rangle$. For this purpose we write the spin correlation as a string of bond variables

$$s_x s_y = \prod_{\langle wz \rangle} b_{\langle wz \rangle}. \tag{2.3.7}$$

The product extends over all bonds $\langle wz \rangle$ connecting the points x and y. The correlation function is then given by

$$\begin{aligned} \langle s_x s_y \rangle &= \frac{1}{Z} \sum_{m=0,1} \prod_{\langle xy \rangle} \sum_{b_{\langle xy \rangle}=\pm 1} \prod_{\langle wz \rangle} b_{\langle wz \rangle} \exp(\beta J b_{\langle xy \rangle}) b_{\langle xy \rangle}^m \\ &= \frac{1}{Z} \{ [2\cosh(\beta J)]^{L-n} [2\sinh(\beta J)]^n \\ &+ [2\sinh(\beta J)]^{L-n} [2\cosh(\beta J)]^n \}. \end{aligned} \tag{2.3.8}$$

Here $n = |x - y|$ is the distance between the points x and y.

The susceptibility is now obtained from

$$\begin{aligned} \chi &= \frac{1}{L} \sum_{x,y} \langle s_x s_y \rangle \\ &= \sum_{n=0}^{L-1} \frac{1}{Z} \{ [2\cosh(\beta J)]^{L-n} [2\sinh(\beta J)]^n \\ &+ [2\sinh(\beta J)]^{L-n} [2\cosh(\beta J)]^n \}. \end{aligned} \tag{2.3.9}$$

Applying the formula for an incomplete geometric series

$$\sum_{n=0}^{L-1} x^n = \frac{1 - x^L}{1 - x} \tag{2.3.10}$$

for $x = \tanh(\beta J)$ as well as $x = \coth(\beta J)$ one obtains

$$\begin{aligned} \chi &= \frac{1}{Z} \{ [2\cosh(\beta J)]^L \frac{1 - \tanh^L(\beta J)}{1 - \tanh(\beta J)} + [2\sinh(\beta J)]^L \frac{1 - \coth^L(\beta J)}{1 - \coth(\beta J)} \} \\ &= \frac{1 - \tanh^L(\beta J)}{1 + \tanh^L(\beta J)} \exp(2\beta J). \end{aligned} \tag{2.3.11}$$

2.4 Exact Results for the 2-d Ising Model

Next we consider the 2-d Ising model. In that case the bond variables around an elementary lattice square with four sites w, x, y, and z satisfy the constraint

$$b_{\langle wx\rangle}b_{\langle xy\rangle}b_{\langle yz\rangle}b_{\langle zw\rangle} = 1. \tag{2.4.1}$$

For each lattice square we introduce a variable $m_\Box$ that implements this constraint via

$$\delta_{b_{\langle wx\rangle}b_{\langle xy\rangle}b_{\langle yz\rangle}b_{\langle zw\rangle},1} = \frac{1}{2}\sum_{m_\Box=0,1}(b_{\langle wx\rangle}b_{\langle xy\rangle}b_{\langle yz\rangle}b_{\langle zw\rangle})^{m_\Box}. \tag{2.4.2}$$

We now introduce the dual lattice with sites $\tilde{x}$ at the centers of the squares $\Box$. The variable $m_\Box$ can then be interpreted as a spin variable,

$$s_{\tilde{x}} = 1 - 2m_\Box = \pm 1, \tag{2.4.3}$$

on the dual lattice. Summing over the bond variable $b_{\langle xy\rangle}$ on the original lattice then induces an interaction between the dual spins $s_{\tilde{x}}$ and $s_{\tilde{y}}$ at the centers of the two squares $\Box_{\tilde{x}}$ and $\Box_{\tilde{y}}$ that share the bond $\langle xy\rangle$. We have

$$\sum_{b_{\langle xy\rangle}=\pm 1}\exp(\beta J b_{\langle xy\rangle})b_{\langle xy\rangle}^{m_{\Box_{\tilde{x}}}+m_{\Box_{\tilde{y}}}} = \exp(-\tilde{\beta}\tilde{h}(s_{\tilde{x}}, s_{\tilde{y}})). \tag{2.4.4}$$

which defines a Hamilton function

$$\tilde{H}[s] = \sum_{\langle\tilde{x}\tilde{y}\rangle}\tilde{h}(s_{\tilde{x}}, s_{\tilde{y}}). \tag{2.4.5}$$

One obtains

$$\begin{aligned}\exp(-\tilde{\beta}\tilde{h}(1,1)) &= \exp(-\tilde{\beta}\tilde{h}(-1,-1)) = 2\cosh(\beta J),\\ \exp(-\tilde{\beta}\tilde{h}(1,-1)) &= \exp(-\tilde{\beta}\tilde{h}(-1,1)) = 2\sinh(\beta J),\end{aligned} \tag{2.4.6}$$

In the original Ising model the ratio of the two Boltzmann factors was

$$\exp(-\beta h(1,-1))/\exp(-\beta h(1,1)) = \exp(-2\beta J). \tag{2.4.7}$$

Similarly, the ratio of the two dual Boltzmann factors is

$$\exp(-\tilde{\beta}\tilde{h}(1,-1))/\exp(-\tilde{\beta}\tilde{h}(1,1)) = \tanh(\beta J) = \exp(-2\tilde{\beta}\tilde{J}). \tag{2.4.8}$$

This equation determines the coupling constant $\tilde{J}$ of a dual Ising model. When the original Ising model is in the high-temperature phase (βJ small) the dual Ising model is in the low-temperature phase ($\tilde{\beta}\tilde{J}$ large) and vice versa. The exact critical temperature $T_c = 1/k_B\beta_c$ of the 2-d Ising model follows from the self-duality condition

$$\tanh(\beta_c J) = \exp(-2\beta_c J), \tag{2.4.9}$$

which again is not identical with the mean field result of eq.(2.2.6).

2.5 Cluster Representation

In this section we will rewrite the Ising model in terms of spin and bond variables. Parallel spins connected by activated bonds form correlated clusters, while spins in different clusters are uncorrelated. The susceptibility can be expressed in terms of the cluster sizes. The cluster representation of the Ising model gives rise to an extremely efficient Monte Carlo algorithm — the so-called cluster algorithm.

We begin by introducing bond variables $b_{\langle xy\rangle} = 0, 1$ which are different from the ones introduced before. An activated bond has $b_{\langle xy\rangle} = 1$ while a deactivated bond has $b_{\langle xy\rangle} = 0$. We can now write

$$\exp(-\beta h(s_x, s_y)) = \sum_{b_{\langle xy\rangle}=0,1} \exp(-\beta h(s_x, s_y, b_{\langle xy\rangle})), \tag{2.5.1}$$

with

$$\begin{aligned}
&\exp(-\beta h(s, s, 1)) = \exp(\beta J) - \exp(-\beta J),\\
&\exp(-\beta h(s, s, 0)) = \exp(-\beta J),\\
&\exp(-\beta h(s, -s, 1)) = 0,\\
&\exp(-\beta h(s, -s, 0)) = \exp(-\beta J),
\end{aligned} \tag{2.5.2}$$

for $s = \pm 1$, such that indeed

$$\begin{aligned}
\exp(-\beta h(s, s)) &= \exp(\beta J)\\
&= \exp(-\beta h(s, s, 1)) + \exp(-\beta h(s, s, 0)),\\
\exp(-\beta h(s, -s)) &= \exp(-\beta J)\\
&= \exp(-\beta h(s, -s, 1)) + \exp(-\beta h(s, -s, 0)).
\end{aligned} \tag{2.5.3}$$

Note that a bond can be activated only if the two connected spins are parallel. It should also be noted that the Boltzmann weight of a deactivated bond is independent of the spin configuration, i.e.

$$\begin{aligned}&\exp(-\beta h(1,1,0)) = \exp(-\beta h(1,-1,0)) = \\ &\exp(-\beta h(-1,1,0)) = \exp(-\beta h(-1,-1,0)) = \exp(-\beta J). \end{aligned} \tag{2.5.4}$$

This implies that the spin configuration decomposes into clusters of parallel spins. The spins connected by activated bonds belong to the same cluster. Spins in the same cluster are hence parallel, while spins in different clusters are uncorrelated. It should be noted that spins completely unconnected to other spins form a cluster by themselves. In this way, each spin belongs to exactly one cluster.

The cluster decomposition of a spin configuration has interesting consequences. In particular, when all spins of a given cluster are flipped, the Boltzmann weight of the spin and bond configuration remains the same. This means that there are 2^{N_C} equally probable configurations, which are obtained by independently flipping all N_C clusters in the configuration. From this fact one can derive a cluster representation of the susceptibility. First, the total magnetization is a sum of cluster magnetizations

$$\mathcal{M}[s] = \sum_x s_x = \sum_C \mathcal{M}_C, \tag{2.5.5}$$

where the cluster magnetization is given by

$$\mathcal{M}_C = \sum_{x \in C} s_x. \tag{2.5.6}$$

In a finite volume the average magnetization always vanishes (even in the broken phase), i.e.

$$\langle \mathcal{M} \rangle = \langle \sum_C \mathcal{M}_C \rangle = 0, \tag{2.5.7}$$

since each cluster of spins can be flipped which leads to a change of sign of the magnetization. Similarly, the susceptibility can be expressed as

$$\begin{aligned} \chi &= \frac{1}{L^d} \langle \mathcal{M}^2 \rangle = \frac{1}{L^d} \langle (\sum_C \mathcal{M}_C)^2 \rangle \\ &= \frac{1}{L^d} \langle \sum_{C_1, C_2} \mathcal{M}_{C_1} \mathcal{M}_{C_2} \rangle = \frac{1}{L^d} \langle \sum_C \mathcal{M}_C^2 \rangle. \end{aligned} \tag{2.5.8}$$

In the last step we have used the fact that two different clusters C_1 and C_2 are uncorrelated, i.e. $\mathcal{M}_{C_1} \mathcal{M}_{C_2}$ averages to zero under cluster flip. Consequently,

only the square of the magnetization of the individual clusters $\mathcal{M}_C^2$ determines the susceptibility. Since all spins of a cluster are parallel, up to a sign the cluster magnetization is given by the cluster size $|C|$, i.e.

$$M_C = \pm|C| = \pm \sum_{x \in C} 1. \tag{2.5.9}$$

Hence, the susceptibility can also be written as

$$\chi = \frac{1}{L^d} \langle \sum_C |C|^2 \rangle. \tag{2.5.10}$$

This shows that the cluster size is directly related to a physical quantity. In this sense the clusters are indeed physical objects.

Chapter 3

The Monte Carlo Method

A powerful numerical technique to solve problems in statistical mechanics is the so-called Monte Carlo method. The idea is to compute expectation values by generating spin configurations numerically. Of course, the partition function is an extremely large sum, such that doing it with numerical brute force is completely hopeless. In the Monte Carlo method predominantly those spin configurations are generated that have the largest contribution to the partition function. In fact, the Boltzmann factor $\exp(-\beta\mathcal{H}[s])$ is used as the probability to generate the spin configuration $[s]$.

3.1 The Concept of a Markov Chain

In a Monte Carlo simulation one generates a sequence of spin configurations

$$[s^{(1)}] \to [s^{(2)}] \to ... \to [s^{(N)}], \tag{3.1.1}$$

which form a so-called Markov chain, by applying an algorithm that turns the configuration $[s^{(i)}]$ into $[s^{(i+1)}]$. The initial configuration $[s^{(1)}]$ is either picked at random or selected otherwise. Ultimately, nothing should depend on this choice. After a (possibly large) number M of Monte Carlo iterations (applications of the algorithm) an equilibrium is reached, and the system has forgotten about the initial configurations. Only the configurations generated after equilibration are used in the actual calculation. To estimate the expectation value of some

observable one averages its values over all configurations of the Monte Carlo sample

$$\langle \mathcal{O} \rangle = \lim_{N \to \infty} \frac{1}{N - M} \sum_{i=M+1}^{N} \mathcal{O}[s^{(i)}]. \tag{3.1.2}$$

In the limit $N \to \infty$ the calculation becomes exact. At finite $N - M$ one makes a calculable statistical error that decreases in proportion to $1/\sqrt{N - M - 1}$. Hence, to increase the numerical accuracy by a factor of two one must run the Monte Carlo algorithm four times as long. The Boltzmann factor $\exp(-\beta \mathcal{H}[s])$ is not explicitly included in the above sum. It is implicitly included, because the configurations in the Markov chain occur with probability $\exp(-\beta \mathcal{H}[s])$.

3.2 Ergodicity and Detailed Balance

To demonstrate that a particular Monte Carlo algorithm converges to the correct equilibrium distribution it is sufficient to show that it is ergodic and obeys detailed balance. Ergodicity means that starting from an arbitrary initial configuration the algorithm can in principle reach any other spin configuration. This condition is obviously necessary, because the correct value for the expectation value can be obtained only if all spin configurations are included. Detailed balance means that

$$\exp(-\beta \mathcal{H}[s]) w[s, s'] = \exp(-\beta \mathcal{H}[s']) w[s', s]. \tag{3.2.1}$$

Here $w[s, s']$ is the transition probability for the algorithm to turn the configuration $[s]$ into $[s']$. A Monte Carlo algorithm is completely characterized by the corresponding $w[s, s']$. Since the algorithm definitely generates a new configuration the proper normalization is

$$\sum_{[s']} w[s, s'] = 1. \tag{3.2.2}$$

When the Monte Carlo algorithm converges to an equilibrium distribution $p[s]$ of spin configurations, this distribution is an eigenvector of $w[s, s']$ with eigenvalue 1

$$\sum_{[s]} p[s] w[s, s'] = p[s']. \tag{3.2.3}$$

Now we want to show that the canonical Boltzmann distribution

$$p[s] = \exp(-\beta \mathcal{H}[s]) \tag{3.2.4}$$

is indeed an eigenvector of $w[s,s']$ if the algorithm obeys detailed balance. We find

$$\begin{aligned}\sum_{[s]} \exp(-\beta\mathcal{H}[s]) w[s,s'] &= \sum_{[s]} \exp(-\beta\mathcal{H}[s']) w[s',s] \\ &= \exp(-\beta\mathcal{H}[s']) \sum_{[s]} w[s',s] \\ &= \exp(-\beta\mathcal{H}[s']). \end{aligned} \tag{3.2.5}$$

Assuming ergodicity one can show that only one eigenvector with eigenvalue 1 exists, and that the equilibrium distribution is therefore unique.

3.3 The Metropolis Algorithm

A simple example of an algorithm that is ergodic and obeys detailed balance is the so-called Metropolis algorithm. In this algorithm a new configuration $[s']$ is randomly chosen based on the old configuration $[s]$. If the energy of the new configuration is smaller than the energy of the old configuration, the new configuration is accepted, i.e.

$$\mathcal{H}[s'] < \mathcal{H}[s] \Rightarrow w[s,s'] = 1. \tag{3.3.1}$$

On the other hand, if the new energy is larger, the new configuration is accepted only with a certain probability, i.e.

$$\mathcal{H}[s'] > \mathcal{H}[s] \Rightarrow w[s,s'] = \exp(-\beta(\mathcal{H}[s'] - \mathcal{H}[s])). \tag{3.3.2}$$

Otherwise the old configuration is kept. This algorithm obeys detailed balance. Let us consider two configurations $[s]$ and $[s']$. We can assume that $\mathcal{H}[s'] < \mathcal{H}[s]$ such that $w[s,s'] = 1$. Then of course, $\mathcal{H}[s] > \mathcal{H}[s']$ such that $w[s',s] = \exp(-\beta(\mathcal{H}[s] - \mathcal{H}[s']))$, and hence

$$\begin{aligned}\exp(-\beta\mathcal{H}[s]) w[s,s'] &= \exp(-\beta\mathcal{H}[s]) \\ &= \exp(-\beta\mathcal{H}[s']) \exp(-\beta(\mathcal{H}[s] - \mathcal{H}[s'])) \\ &= \exp(-\beta\mathcal{H}[s']) w[s',s]. \end{aligned} \tag{3.3.3}$$

We still need to specify how a new configuration is proposed. In the Ising model one visits the spins one by one and proposes to flip them. The resulting

change of the energy is calculated by investigating the neighboring spins. Then following the Metropolis algorithm, it is decided if a given spin is flipped or not. When all spins on the lattice have been updated in this way one has completed one Metropolis sweep. It is obvious that any spin configuration can, at least in principle, be reached in this way, i.e. the Metropolis algorithm is indeed ergodic. A typical Monte Carlo simulation consists of a large number of sweeps, say 1 million, for example.

3.4 Error Analysis

Since any practical Monte Carlo simulation has a finite length, the results are not exact but are affected by statistical errors. Hence, an important part of every Monte Carlo calculation is the error analysis. An ideal Monte Carlo algorithm (which doesn't exist in practice) would generate a Markov chain of statistically independent configurations. If the Monte Carlo data for an observable $\mathcal{O}$ are Gaussian distributed, the standard deviation from their average (i.e. their statistical error) is given by

$$\Delta\mathcal{O} = \frac{1}{\sqrt{N-M-1}}\langle(\mathcal{O}-\langle\mathcal{O}\rangle)^2\rangle = \frac{1}{\sqrt{N-M-1}}(\langle\mathcal{O}^2\rangle - \langle\mathcal{O}\rangle^2). \qquad (3.4.1)$$

In order to reduce the statistical error by a factor of two, the number of independent equilibrated configurations $N-M$ must hence be increased by a factor of four.

Practical Monte Carlo algorithms (like the Metropolis algorithm) are not ideal, i.e. they do not generate statistically independent configurations. In particular, the Metropolis algorithm is rather simple, but not very efficient. Since the new configuration is generated from the previous configuration in the Markov chain, subsequent configurations are correlated. This implies that the actual statistical error is larger than the above naive estimate of the standard deviation. In order to detect the autocorrelation of the Monte Carlo data it is useful to bin these data. For this purpose one averages a number N_b of subsequent measurements and treats this average as a statistically independent result. One then computes the standard deviation based on the $(N-M)/N_b$ statistically independent averages. Of course, if the bin size N_b is too small, the averages are still correlated and the corresponding standard deviation still underestimates the true statistical

error. When one increases the bin size N_b, the corresponding standard deviation increases until subsequent bin averages are indeed statistically independent. Once the standard deviation has reached a plateau (by increasing N_b), one has obtained a reliable estimate of the true statistical error.

In order to estimate the number τ of Monte Carlo iterations that separate statistically independent spin configuration, it is also useful to determine the autocorrelation function of some observable $\mathcal{O}$

$$\langle \mathcal{O}^{(i)} \mathcal{O}^{(i+t)} \rangle = \lim_{N \to \infty} \frac{1}{N - M - t} \sum_{i=M+1}^{N-t} \mathcal{O}[s^{(i)}] \mathcal{O}[s^{(i+t)}] \propto \exp(-t/\tau). \tag{3.4.2}$$

The autocorrelation time τ of the Metropolis algorithm actually increases when one approaches a second order phase transition. At a second order phase transition the correlation length ξ diverges. One finds so-called critical slowing down

$$\tau \propto \xi^z, \tag{3.4.3}$$

where z is a dynamical critical exponent characterizing the efficiency of a Monte Carlo algorithm. For the Metropolis algorithm one finds $z \approx 2$, which leads to a very bad critical slowing down behavior. This is a good motivation to turn to the much more efficient cluster algorithms which can reach $z \approx 0$.

3.5 The Swendsen-Wang Cluster Algorithm

The Swendsen-Wang cluster algorithm is a Monte Carlo method based on the cluster representation of the Ising model, i.e. it operates on both spins and bonds. First, an initial spin configuration is selected. Then each bond is activated or deactivated, depending on the orientation of the adjacent spins. If the two spins connected by the bond are antiparallel the bond is deactivated. On the other hand, if the two spins are parallel the bond can be activated or deactivated. The corresponding Boltzmann weights are given by

$$\begin{aligned} \exp(-\beta h(s, s, 1)) &= \exp(\beta J) - \exp(-\beta J), \\ \exp(-\beta h(s, s, 0)) &= \exp(-\beta J), \end{aligned} \tag{3.5.1}$$

where $s = \pm 1$. Hence, the probability for activating the bond is

$$p = \frac{\exp(-\beta h(s, s, 1))}{\exp(-\beta h(s, s, 1)) + \exp(-\beta h(s, s, 0))} = 1 - \exp(-2\beta J). \tag{3.5.2}$$

Once each bond is activated or deactivated, one identifies the clusters of sites connected by activated bonds. By construction all spins in a common cluster are parallel. Each cluster is flipped (i.e. all spins in the cluster change sign) with 50 percent probability. This completes one sweep of the cluster algorithm. Then the procedure is repeated, i.e. the bonds are updated again.

As a benefit of the cluster algorithm we can make use of the cluster representation of the susceptibility

$$\chi = \frac{1}{L^d} \langle \sum_C |C|^2 \rangle \tag{3.5.3}$$

in order to obtain a so-called improved estimator. Instead of measuring just $\mathcal{M}[s]^2$ for the given spin configuration, we sum the squares of all cluster sizes $|C|$. Effectively, this increases the statistics by a factor 2^{N_c}, where N_C is the number of clusters in the configuration.

Let us first show that the cluster algorithm is ergodic. There is a finite (although perhaps very small) probability that no bonds are activated. Then each spin forms its own cluster. By flipping these individual spin clusters one can obviously reach any possible spin configuration.

We still need to show that the cluster algorithm obeys detailed balance, i.e.

$$\exp(-\beta \mathcal{H}[s]) w[s, s'] = \exp(-\beta \mathcal{H}[s']) w[s', s]. \tag{3.5.4}$$

It is sufficient to consider just one pair of neighboring spins. If the two spins are antiparallel they necessarily belong to different clusters. After cluster flip they will be parallel with 50 percent probability. In the next sweep the bond between them will then be activated with probability p and deactivated with probability $(1-p)$. The probability to turn back into the original antiparallel configuration is then $\frac{1}{2}(1-p)$. The corresponding detailed balance relation then takes the form

$$\begin{aligned} &\exp(-\beta h(s,-s))\frac{1}{2} = \exp(-\beta h(s,s))\frac{1}{2}(1-p) \;\Rightarrow \\ &\exp(-\beta J)\frac{1}{2} = \exp(\beta J)\frac{1}{2}\exp(-2\beta J), \end{aligned} \tag{3.5.5}$$

which is indeed satisfied. With the other 50 percent probability the originally antiparallel spins will remain antiparallel. In that case, the bond between them cannot be activated and thus with 50 percent probability we return to the original configuration. Detailed balance is then trivially satisfied. Finally, let us assume

that the two spins are originally parallel. Then we need to distinguish between two cases. First, we assume that the two spins are already indirectly connected through other activated bonds. In that case, it is irrelevant if the direct bond between them is activated or not. The two spins remain parallel and detailed balance is trivially satisfied. Next, we assume that the two parallel spins are not indirectly connected by activated bonds. Then the direct bond between them is activated with probability p and the spins remain parallel. With probability $(1-p)$ the bond is deactivated and the spins remain parallel only with 50 percent probability. Again, detailed balance is then trivially satisfied. With the other 50 percent probability the two spins will become antiparallel. Then they cannot be bound by an activated bond and they return into the parallel configuration with 50 percent probability. The detailed balance relation then again takes the form

$$\begin{aligned} &\exp(-\beta h(s,s))\frac{1}{2}(1-p) = \exp(-\beta h(s,-s))\frac{1}{2} \Rightarrow \\ &\exp(\beta J)\frac{1}{2}\exp(-2\beta J) = \exp(-\beta J)\frac{1}{2}. \end{aligned} \tag{3.5.6}$$

Thus, in all cases detailed balance is indeed satisfied.

3.6 The Wolff Cluster Algorithm

The Wolff cluster algorithm is an interesting (and sometimes even more efficient) variant of the Swendsen-Wang algorithm. The Swendsen-Wang algorithm is a multi-cluster algorithm, i.e. all clusters in a configuration are identified and on average half of them are flipped. The Wolff cluster algorithm, on the other hand, is a single-cluster algorithm, i.e. a site is selected at random and only the one cluster attached to this site is identified in this configuration. In contrast to the Swendsen-Wang algorithm, the single cluster is then flipped with 100 percent probability. In this case, no effort is wasted for identifying clusters which are then not flipped. After the single cluster is flipped, a new bond configuration is generated and the whole procedure is repeated.

As for the Swendsen-Wang algorithm, for the Wolff cluster algorithm one can also construct an improved estimator for the susceptibility. While in the multi-cluster algorithm all clusters contribute $|C|^2$ to the susceptibility, in the single-cluster algorithm the cluster is selected with a probability $|C|/L^d$ (proportional to its size $|C|$). Hence, bigger clusters are selected more frequently than smaller

ones and one must correct for this bias. Hence, in the single-cluster algorithm the improved estimator for the susceptibility takes the form

$$\chi = \frac{1}{L^d}\langle |C|^2 \frac{L^d}{|C|}\rangle = \langle |C|\rangle. \tag{3.6.1}$$

Chapter 4

The Complex Action Problem in Dense QCD

In this chapter we investigate the complex action problem in the Potts model approximation to dense QCD. While it is presently impossible to simulate QCD at non-zero quark chemical potential due to a severe complex action problem, in the Potts model approximation the complex action problem can be solved using a cluster algorithm.

4.1 QCD with Heavy Quarks and the 3-d 3-State Potts Model

The partition function for a pure $SU(3)$ Yang-Mills theory is given by

$$Z = \int \mathcal{D}A \ \exp(-S[A]), \tag{4.1.1}$$

where

$$S[A] = \int_0^\beta dt \int d^3x \ \frac{1}{2g^2} \mathrm{Tr}[F_{\mu\nu}F_{\mu\nu}], \tag{4.1.2}$$

is the Euclidean action for the gluons and β is the inverse temperature. The action is invariant under gauge transformations

$$g(\vec{x}, 0) = g(\vec{x}, \beta)z, \tag{4.1.3}$$

that are periodic in Euclidean time up to an element z of the center $\mathbf{Z}(3) = \{\exp(2\pi in/3), n = 1, 2, 3\}$ of the non-Abelian gauge group. In the presence of a single external heavy quark of bare mass M at an undetermined position $\vec{x}$ the partition function turns into

$$Z_Q = \int \mathcal{D}A \ \Phi[A] \exp(-S[A]) \exp(-\beta M), \tag{4.1.4}$$

where

$$\Phi[A] = \int d^3x \ \mathrm{Tr}[\mathcal{P} \exp(-\int_0^\beta dt \ A_4(\vec{x}, t))], \tag{4.1.5}$$

is the spatial integral of the Polyakov loop. Ultimately, the mass M will be sent to infinity. Note that while the center transformation of eq.(4.1.3) leaves the pure gluon action $S[{}^gA] = S[A]$ invariant, the Polyakov loop transforms into

$$\Phi[{}^gA] = z\Phi[A]. \tag{4.1.6}$$

This shows that in the presence of the external quark, the $\mathbf{Z}(3)$ symmetry is explicitly broken. The partition function for a system of gluons in the presence of a single heavy anti-quark is given by

$$Z_{\bar{Q}} = \int \mathcal{D}A \ \Phi[A]^* \exp(-S[A]) \exp(-\beta M), \tag{4.1.7}$$

where $*$ denotes complex conjugation. Let us now consider a system of gluons in a background of n static quarks and $\bar{n}$ static anti-quarks. The partition function then takes the form

$$Z_{n,\bar{n}} = \int \mathcal{D}A \ \frac{1}{n!}\Phi[A]^n \frac{1}{\bar{n}!}(\Phi[A]^*)^{\bar{n}} \exp(-S[A]) \exp(-\beta M(n + \bar{n})). \tag{4.1.8}$$

The factors $n!$ and $\bar{n}!$ appear because quarks as well as anti-quarks are indistinguishable. Introducing the quark chemical potential μ that couples to the baryon number $B = (n - \bar{n})/3$ we obtain the grand canonical partition function

$$\begin{aligned} Z(\mu) &= \sum_{n,\bar{n}} Z_{n,\bar{n}} \exp(\beta\mu(n - \bar{n})) \\ &= \sum_{n,\bar{n}} \int \mathcal{D}A \ \frac{1}{n!}\Phi[A]^n \frac{1}{\bar{n}!}(\Phi[A]^*)^{\bar{n}} \\ &\times \ \exp(-S[A] - \beta n(M - \mu) - \beta\bar{n}(M + \mu)) \\ &= \int \mathcal{D}A \ \exp(-S[A]) \\ &\times \ \exp(\exp(-\beta(M - \mu))\Phi[A] + \exp(-\beta(M + \mu))\Phi[A]^*). \end{aligned} \tag{4.1.9}$$

As expected, the presence of quarks and anti-quarks leads to an explicit breaking of the $\mathbb{Z}(3)$ center symmetry. Furthermore, in the presence of a non-zero chemical potential the effective action for the gluons is complex. Note that in the $SU(2)$ case the action remains real because then the Polyakov loop itself is real, i.e. $\Phi[A]^* = \Phi[A]$. The action becomes real even in the $SU(3)$ case if μ is purely imaginary. Furthermore, one can see that the chemical potential explicitly breaks the charge conjugation symmetry that replaces $\Phi[A]$ by $\Phi[A]^*$. In fact, under charge conjugation the action turns into its complex conjugate. We have assumed that the quarks are static. Hence, to be consistent we must consider the limit $M \to \infty$. In order to obtain a non-trivial result, we simultaneously take the limit $\mu \to \infty$ such that $M - \mu$ remains finite. The partition function then simplifies to

$$Z(\mu) = \int \mathcal{D}A \ \exp(-S[A] + \exp(-\beta(M-\mu))\Phi[A]). \tag{4.1.10}$$

Up to this point we have treated QCD consistently in the static quark limit. The resulting effective action for the gluons is complex and we presently don't know how to simulate it efficiently. For that reason we now replace the gluon system by a simple 3-d lattice 3-state Potts model. The Potts spins $\Phi_x \in \mathbb{Z}(3)$ replace the original Polyakov loop variables and the partition function turns into

$$Z(h) = \int \mathcal{D}\Phi \ \exp(-S[\Phi] + h\sum_x \Phi_x), \tag{4.1.11}$$

where h replaces $\exp(-\beta(M-\mu))$. Note that the Potts model action is still complex. In principle, one can imagine to integrate out all QCD degrees of freedom except for the $\mathbb{Z}(3)$ phase of the Polyakov loop and thus derive an effective Potts model action directly from QCD. In practice this is impossible, except in the strong coupling limit. For simplicity, we therefore replace the pure gluon action $S[A]$ by a standard nearest-neighbor Potts model interaction

$$S[\Phi] = -\kappa \sum_{x,i} \delta_{\Phi_x, \Phi_{x+\hat{i}}}. \tag{4.1.12}$$

The coupling constant κ is not directly related to the parameters of QCD. Still, a large value of κ corresponds qualitatively to the high-temperature deconfined phase, while small κ values correspond to the confined phase. As mentioned in the introduction, the Potts model also retains the general features of the QCD phase diagram. At $h = 0$ (M infinite, μ finite) there is a first-order phase transition

as a function of κ, between the disordered (confined) phase that respects the $\mathbf{Z}(3)$ symmetry and the ordered (deconfined) phase that spontaneously breaks it. An order parameter for this transition is $\langle\Phi\rangle$. As h rises from zero, the chemical potential term explicitly breaks the $\mathbf{Z}(3)$ symmetry, the phase transition weakens, and then ends at a critical point. Correspondingly, in heavy-quark QCD the quarks begin to contribute to the partition function when μ gets close to M, and there is no longer an order parameter for deconfinement. The deconfining phase transition terminates at a critical endpoint.

4.2 The General Nature of the Complex Action Problem

When the action is complex the resulting Boltzmann factor cannot be interpreted as a probability and hence standard importance sampling techniques fail. When one uses just the absolute value of the Boltzmann factor for importance sampling and includes its complex phase in measured observables O, expectation values take the form

$$\begin{aligned}\langle O\rangle &= \frac{1}{Z}\int \mathcal{D}\Phi\ O[\Phi]\exp(-S[\Phi]+h\sum_x \Phi_x)\\ &= \frac{\langle O\exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R}{\langle\exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R}.\end{aligned} \tag{4.2.1}$$

The subscript R refers to a modified ensemble with a real action described by the partition function

$$Z_R = \int \mathcal{D}\Phi\ \ \exp(-S[\Phi]+h\sum_x \mathrm{Re}\Phi). \tag{4.2.2}$$

By definition we have

$$\begin{aligned}\langle\exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R &= \frac{1}{Z_R}\int \mathcal{D}\Phi\ \ \exp(ih\sum_x \mathrm{Im}\Phi_x)\exp(-S[\Phi]+h\sum_x \mathrm{Re}\Phi)\\ &= \frac{Z}{Z_R}\approx\exp(-V(f-f_R)),\end{aligned} \tag{4.2.3}$$

where f and f_R are the free energy densities of the original complex and the modified real action systems, respectively, and V is the spatial volume. Hence, the

denominator in eq.(4.2.1) becomes exponentially small as one increases the volume. The same is true for the numerator, because $\langle O \rangle$ itself is not exponentially large in V.

Although, in principle, simulating the modified ensemble is correct, in practice this method fails for large volumes. The reason is that observables are obtained as ratios of exponentially small numerators and denominators which are themselves averages of quantities of order one. This leads to very severe cancellations and requires an exponentially large statistics in order to obtain accurate results. To see this, we estimate the relative statistical error in the determination of the average phase of the Boltzmann factor $\exp(ih\sum_x \mathrm{Im}\Phi_x)$. Since $\langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R = Z/Z_R$ the average itself is real. When one generates N statistically independent field configurations in a Monte Carlo simulation, the resulting error to signal ratio is given by

$$\begin{aligned}
\frac{\Delta \exp(ih\sum_x \mathrm{Im}\Phi_x)}{\langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R} &= \frac{\sqrt{\langle |\exp(ih\sum_x \mathrm{Im}\Phi_x) - \langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R|^2\rangle_R}}{\sqrt{N}\langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R} \\
&= \frac{\sqrt{1 - \langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R^2}}{\sqrt{N}\langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R} \\
&\approx \frac{\exp(V(f-f_R))}{\sqrt{N}}. \qquad (4.2.4)
\end{aligned}$$

For large V we have used $\langle \exp(ih\sum_x \mathrm{Im}\Phi_x)\rangle_R \ll 1$ as implied by eq.(4.2.3). Consequently, in order to obtain an acceptable error to signal ratio one must generate at least $N \approx \exp(2V(f-f_R))$ configurations. For large volumes this is impossible in practice.

4.3 Cluster Algorithm for the Potts Model

Let us now outline the ideas that underlie the cluster algorithm that we use to solve the complex action problem. It is based on the original Swendsen-Wang cluster algorithm for the Potts model without chemical potential. In fact, in the limit $h = 0$ our algorithm reduces to that algorithm. The Swendsen-Wang cluster algorithm decomposes the lattice into independent clusters of connected sites. Each spin belongs to exactly one cluster and all spins within a cluster are assigned the same random $\mathbf{Z}(3)$ element. In this section, we construct an

improved estimator for the h-dependent part $\exp(h\sum_x \Phi_x)$ of the Boltzmann factor by analytically averaging it over all configurations related to each other by cluster flips. Although, for an individual configuration $\exp(h\sum_x \Phi_x)$ is in general complex, its improved estimator is always real and positive and can thus be used for importance sampling. This completely solves the complex action problem.

Let us first describe the original Swendsen-Wang algorithm for $h = 0$. In this method one introduces variables $b = 0, 1$ for each bond connecting neighboring lattice sites x and $y = x + \hat{i}$ and one writes the nearest neighbor Boltzmann factor as

$$\exp(\kappa \delta_{\Phi_x,\Phi_y}) = \sum_{b=0,1} [\delta_{b,1}\delta_{\Phi_x,\Phi_y}(e^\kappa - 1) + \delta_{b,0}]. \tag{4.3.1}$$

In the enlarged configuration space of spin and bond variables, the bond variables impose constraints between the spin variables. When a bond is put (i.e. when $b = 1$), the spin Boltzmann factor is $\delta_{\Phi_x,\Phi_y}(e^\kappa - 1)$ and hence the spin variables Φ_x and Φ_y at the two ends of the bond must be identical. On the other hand, when the bond is not put ($b = 0$), the spin Boltzmann factor is 1 and thus the variables Φ_x and Φ_y fluctuate independently. The spin variables, in turn, determine the probability to put a bond. When the spins Φ_x and Φ_y are different, the bond Boltzmann factor is $\delta_{b,0}$ and thus the bond is not put. On the other hand, when Φ_x and Φ_y are the same, the bond Boltzmann factor is $[\delta_{b,1}(e^\kappa - 1) + \delta_{b,0}]$. Consequently, a bond between parallel spins is put with probability $p = 1 - e^{-\kappa}$. Note that for $\kappa = 0$ no bonds are put, while for $\kappa = \infty$ parallel spins are always connected by a bond.

The Swendsen-Wang cluster algorithm updates bond and spin variables in alternating order. First, for a given spin configuration, bonds are put with probability p between parallel neighboring spins. No bonds are put between non-parallel spins. Then the spins are updated according to the constraints represented by the resulting bond configuration. Spins connected by bonds must remain parallel, while spins not connected by bonds fluctuate independently. Hence, to update the spins, one must identify clusters, i.e. sets of spins that are connected by bonds. All spins in a cluster are parallel and are assigned the same random $\mathbb{Z}(3)$ element in the spin update. All spins belong to exactly one cluster. It should be noted that a cluster may consist of a single spin. A configuration consisting of N_C clusters can be viewed as a member of a sub-ensemble of 3^{N_C} equally probable configurations which result by assigning $\mathbb{Z}(3)$ elements to the various clusters in all possible ways. As was already pointed out by Swendsen and Wang, one can construct improved estimators for various physical quantities by averaging

analytically over all 3^{N_C} configurations in a sub-ensemble. Since the number of clusters is proportional to the volume, this effectively increases the statistics by a factor that is exponentially large in V.

Next, let us construct an improved estimator for the h-dependent part $\exp(h\sum_x \Phi_x)$ of the Boltzmann factor. Although for an individual configuration this term is in general complex, its average over a sub-ensemble of 3^{N_C} configurations is always real and positive. This results from the following observations. The h-dependent part of the Boltzmann factor is a product of cluster contributions

$$\exp(h\sum_x \Phi_x) = \prod_C \exp(h\sum_{x\in C} \Phi_x). \tag{4.3.2}$$

Since the clusters are independent, the sub-ensemble average is a product

$$\langle \exp(h\sum_x \Phi_x)\rangle_{3^{N_C}} = \prod_C \langle \exp(h\sum_{x\in C} \Phi_x)\rangle_3. \tag{4.3.3}$$

of 3-state averages for the individual clusters

$$\begin{aligned} \langle \exp(h\sum_{x\in C} \Phi_x)\rangle_3 &= \frac{1}{3}\sum_{\Phi\in \mathbf{Z}(3)} \exp(h|C|\Phi) \\ &= \frac{1}{3}[\exp(h|C|) + 2\exp(-h|C|/2)\cos(\sqrt{3}h|C|/2)] \\ &= W(C), \end{aligned} \tag{4.3.4}$$

which defines a weight $W(C)$ for each cluster. We have used the fact that all spins Φ_x in a given cluster C take the same value $\Phi \in \mathbf{Z}(3)$ so that $\sum_{x\in C}\Phi_x = |C|\Phi$ where $|C| = \sum_{x\in C} 1$ is the cluster size. It is easy to show that the expression in eq.(4.3.4) is always positive and can hence be used for importance sampling. This is crucial for a complete solution of the complex action problem.

For a given bond configuration one can integrate out the spin variables and one obtains

$$\begin{aligned} Z &= \int \mathcal{D}b\ (e^\kappa - 1)^{N_b}\ 3^{N_C} \prod_C W(C) \\ &= \int \mathcal{D}b\ (e^\kappa - 1)^{N_b} \prod_C [\exp(h|C|) + 2\exp(-h|C|/2)\cos(\sqrt{3}h|C|/2)]. \end{aligned} \tag{4.3.5}$$

Here N_b is the number of bonds that are put (i.e. have $b = 1$). The factor 3^{N_C} represents the number of allowed spin configurations for a given bond configuration

and the factors $W(C)$ come from the improved estimator. The effective action for the bond variables depends only on the sizes $|C|$ of the clusters corresponding to a given bond configuration. Note that the factor $1/3$ per cluster in eq.(4.3.4) cancels against the factor 3^{N_C}.

Our algorithm directly updates the partition function of eq.(4.3.5), i.e. it only operates on the bond variables while the spins are already integrated out analytically. The bond variables that define the clusters are updated with a local algorithm. A bond whose value does not affect the cluster sizes is put with probability $p = 1 - e^{-\kappa}$. This happens when the two sites at its ends belong to the same cluster because they are connected indirectly through other bonds. A bond whose value affects the cluster sizes is put with a probability that depends on the sizes of the corresponding clusters. When the bond is not put ($b = 0$), its endpoints x and y belong to two different clusters C_1 and C_2 of sizes $|C_1|$ and $|C_2|$ and the corresponding Boltzmann weight is $3^2 W(C_1)W(C_2)$. On the other hand, when the bond is put ($b = 1$), its endpoints belong to the combined cluster $C_1 \cup C_2$ of size $|C_1| + |C_2|$. In that case, the Boltzmann weight is $3W(C_1 \cup C_2)(e^\kappa - 1)$. Hence, the bond is put with probability

$$q = \frac{W(C_1 \cup C_2)(e^\kappa - 1)}{3W(C_1)W(C_2) + W(C_1 \cup C_2)(e^\kappa - 1)}. \tag{4.3.6}$$

4.4 Improved Estimator for the Polyakov Loop

In order to measure physical observables, it is crucial to construct improved estimators for them as well. Here we construct improved estimators for the Polyakov loop Φ_x, whose expectation value

$$\langle \Phi_x \rangle = \exp(-\beta F_Q), \tag{4.4.1}$$

determines the free energy F_Q of a quark. The improved estimator for the Polyakov loop is given by the sub-ensemble average

$$\langle \Phi_x \exp(h \sum_z \Phi_z) \rangle_{3^{N_C}} = \frac{1}{3} \sum_{\Phi \in \mathbf{Z}(3)} \Phi \exp(h|C_x|\Phi) \prod_{C \neq C_x} W(C), \tag{4.4.2}$$

where C_x is the cluster that contains the point x. Hence, we obtain

$$\langle \Phi_x \rangle = \frac{1}{Z(h)} \int \mathcal{D}b \, \frac{1}{3W(C_x)} \sum_{\Phi \in \mathbf{Z}(3)} \Phi \exp(h|C_x|\Phi)(e^\kappa - 1)^{N_b} \, 3^{N_C} \prod_C W(C), \tag{4.4.3}$$

i.e. after integrating out the spin variables, the Polyakov loop is represented by

$$\Phi_x = \frac{1}{3W(C_x)} \sum_{\Phi \in \mathbf{Z}(3)} \Phi \exp(h|C_x|\Phi). \tag{4.4.4}$$

4.5 Severity of the Complex Action Problem

In order to estimate the severity of the complex action problem, we like to determine the expectation value of the complex phase of the Boltzmann factor in the modified real action ensemble

$$\langle \exp(ih \sum_x \mathrm{Im}\Phi_x) \rangle_R = \frac{Z}{Z_R}. \tag{4.5.1}$$

Rather than implementing this directly in a simulation that uses the absolute value of the Boltzmann factor for importance sampling, one can measure Z_R/Z with the cluster algorithm. In fact, an improved estimator for this quantity is given by $\prod_C W_R(C)/W(C)$, where

$$W_R(C) = \langle \exp(h \sum_{x \in C} \mathrm{Re}\Phi_x) \rangle_3 = \frac{1}{3}[\exp(h|C|) + 2\exp(-h|C|/2)], \tag{4.5.2}$$

is the weight that replaces $W(C)$ in the real action ensemble. Alternatively, one can construct a cluster algorithm that simulates the real action ensemble. In that case, one needs to measure $\prod_C W(C)/W_R(C)$ in order to obtain Z/Z_R. Indeed, one finds an exponentially small signal, as expected from eq.(4.2.3). It should be noted that the complex action problem is most severe for intermediate values of h. While it is obvious that there is no complex action problem at $h = 0$, it is perhaps less obvious that there is also no problem for large h. This is because

$$\frac{W_R(C)}{W(C)} = \frac{\exp(h|C|) + 2\exp(-h|C|/2)}{\exp(h|C|) + 2\exp(-h|C|/2)\cos(\sqrt{3}h|C|/2)} \tag{4.5.3}$$

approaches 1 in the limit $h \to \infty$ so that $\langle \exp(ih \sum_x \mathrm{Im}\Phi_x)_R \rangle_R = Z_R/Z \to 1$.

4.6 The Phase Diagram

Figure 1.1 shows the phase diagram of the model defined by eq.(4.1.11). For $h = 0$ our model reduces to the standard 3-d 3-state Potts model which has

been studied extensively in Monte-Carlo simulations. The model is known to have a weak first order phase transition. The value of the coupling κ where the phase transition occurs (point T in fig. 1) has been determined with very high precision. The phase transition was found to occur at $\kappa_T = 0.550565(10)$. Above this value the $\mathbf{Z}(3)$ symmetry is spontaneously broken, i.e. for $\kappa > \kappa_T$ three distinct deconfined phases coexist. When we switch on the parameter h, the $\mathbf{Z}(3)$ symmetry gets explicitly broken. Positive values of h favor the deconfined phase with a real value of $\langle\Phi\rangle$. Hence, the line $\kappa > \kappa_T$ at $h = 0$ is a line of first order phase transitions which cannot terminate in the deconfinement transition at the point T. In fact, T is a triple point because two other first order transition lines emerge from it. For $h > 0$ a line of first order transitions extends into the (h, κ)-plane and terminates in a critical endpoint (E in fig. 1). Negative values of h favor the two deconfined phases with complex values of $\langle\Phi\rangle$. Negative h are unphysical in the QCD interpretation of the Potts model because h represents $\exp(-\beta(M - \mu))$ in QCD. Still, the Potts model at $h < 0$ makes perfect sense as a statistical mechanics system (unrelated to QCD) and it has another first order transition line emerging from the point T. Interestingly, with our method the sign problem can only be solved for $h \geq 0$ since otherwise the improved estimator of eq.(4.3.4) is not necessarily positive.

The line of first order phase transitions $h_t(\kappa)$ is determined by the condition that the free energy densities of the confined and deconfined phases are equal, i.e. $f_c(h_t(\kappa), \kappa) = f_d(h_t(\kappa), \kappa)$. Close to the point $T = (0, \kappa_T)$ the free energy density of the confined phase is given by

$$f_c(h, \kappa) = f_{c,T} + e_{c,T}(\kappa - \kappa_T), \tag{4.6.1}$$

where $f_{c,T} = f_c(0, \kappa_T)$ and $e_{c,T} = df_c/d\kappa(0, \kappa_T)$ is the energy density of the confined phase at the point T. Note that to leading order $f_c(h, \kappa)$ is independent of h because $\langle\Phi\rangle = 0$ in the confined phase at $h = 0$. On the other hand, for the deconfined phase one obtains

$$f_d(h, \kappa) = f_{d,T} + e_{d,T}(\kappa - \kappa_T) - h\langle\Phi\rangle_T, \tag{4.6.2}$$

where $\langle\Phi\rangle_T$ is the value of the Polyakov loop at the point T in the deconfined phase that is favored at $h > 0$. Using the condition $f_{c,T} = f_{d,T}$ for the deconfinement phase transition at $h = 0$, one finds

$$h_t(\kappa) = \frac{e_{d,T} - e_{c,T}}{\langle\Phi\rangle_T}(\kappa - \kappa_T) = a(\kappa_T - \kappa). \tag{4.6.3}$$

Our data are consistent with the transition line $h_t(\kappa)$ being a straight line. Fitting the values of $h_t(\kappa)$ obtained from a finite size scaling analysis yields $a = 0.447(6)$ with $\chi^2/\text{d.o.f.} = 0.80$. Similarly, one can determine the angle at which the third transition line leaves the point T in the direction of negative h.

4.7 Conclusions

We have used a cluster algorithm to solve the notorious complex action problem in the Potts model for QCD with heavy quarks at large chemical potential. In fact, a simple analytic construction of an improved estimator leads to an exponential reduction of the required statistics. Since the improved estimator is real and positive, importance sampling techniques that fail for complex actions then become applicable. This has allowed us to study the phase diagram of the Potts model at non-zero density in great detail. In particular, we were able to locate the critical endpoint of the first order deconfinement phase transition rather precisely. Our algorithm belongs to the class of meron-cluster algorithms that has recently been used to solve a large variety of sign and complex action problems. Of course, the ultimate goal is to construct a similar algorithm for QCD at non-zero chemical potential and investigate the phase structure of QCD at $\mu \neq 0$ from first principles. The complex action problem in full QCD is more complicated than the one in the Potts model because it interferes with a fermion sign problem. So far, meron-cluster algorithms have led to solutions of fermion sign problems as well as complex action problems, but have not yet solved both problems simultaneously.

Chapter 5

Fermion Sign Problem and Meron-Cluster Algorithm

In this chapter we investigate the fermion sign problem in a simple model of staggered fermions. The meron-cluster algorithm is a method that allows us to completely solve the sign problem in this system.

5.1 A Staggered Fermion Model

Let us consider staggered fermions hopping on a 3-dimensional cubic spatial lattice with L^3 sites x (L even) and with anti-periodic spatial boundary conditions. The fermions are described by creation and annihilation operators Ψ_x^+ and Ψ_x with standard anti-commutation relations

$$\{\Psi_x^+, \Psi_y^+\} = \{\Psi_x, \Psi_y\} = 0, \{\Psi_x^+, \Psi_y\} = \delta_{xy}. \tag{5.1.1}$$

The staggered fermion Hamilton operator takes the form

$$H = \sum_{x,i} h_{x,i} + m \sum_x (-1)^{x_1+x_2+x_3} \Psi_x^+ \Psi_x, \tag{5.1.2}$$

that is a sum of nearest-neighbor couplings $h_{x,i}$ and a mass term $m\overline{\Psi}\Psi$. In the following we work directly in the chiral limit, $m = 0$, and only use $\overline{\Psi}\Psi$ as an

observable. The term $h_{x,i}$ couples the fermion operators at the lattice sites x and $x+\hat{i}$, where $\hat{i}$ is a unit-vector in the i-direction, and

$$h_{x,i} = \frac{1}{2}\eta_{x,i}(\Psi_x^+ \Psi_{x+\hat{i}} + \Psi_{x+\hat{i}}^+ \Psi_x) + G(\Psi_x^+ \Psi_x - \frac{1}{2})(\Psi_{x+\hat{i}}^+ \Psi_{x+\hat{i}} - \frac{1}{2}). \qquad (5.1.3)$$

Here $\eta_{x,1} = 1$, $\eta_{x,2} = (-1)^{x_1}$ and $\eta_{x,3} = (-1)^{x_1+x_2}$ are the standard staggered fermion sign factors, and G is a four-fermion coupling constant. The system has a conserved fermion number

$$N = \sum_x \Psi_x^+ \Psi_x, \qquad (5.1.4)$$

because $[H, N] = 0$. Besides the $U(1)$ fermion number symmetry, the model has a $\mathbf{Z}(2)$ chiral symmetry, which simply shifts Ψ_x and Ψ_x^+ by one lattice spacing in all three directions. This changes the sign of $\overline{\Psi}\Psi$ but leaves the $m = 0$ Hamiltonian invariant. There are also other $\mathbf{Z}(2)$ symmetries which correspond to discrete flavor transformations.

To construct a path integral for the partition function, we decompose the Hamilton operator into six terms $H = H_1 + H_2 + ... + H_6$, with

$$H_i = \sum_{\substack{x=(x_1,x_2,x_3) \\ x_i even}} h_{x,i},\ H_{i+3} = \sum_{\substack{x=(x_1,x_2,x_3) \\ x_i odd}} h_{x,i}. \qquad (5.1.5)$$

The individual contributions to a given H_i commute with each other, but two different H_i do not commute. Using the Suzuki-Trotter formula we express the fermionic partition function at inverse temperature β as

$$Z_f = \mathrm{Tr}[\exp(-\beta H)] = \lim_{M\to\infty} \mathrm{Tr}[\exp(-\epsilon H_1)\exp(-\epsilon H_2)...\exp(-\epsilon H_6)]^M. \qquad (5.1.6)$$

We have introduced $6M$ Euclidean time slices with $\epsilon = \beta/M$ being the lattice spacing in the Euclidean time direction. Following Jordan and Wigner we represent the fermion operators by Pauli matrices

$$\Psi_x^+ = \sigma_1^3\sigma_2^3...\sigma_{l-1}^3\sigma_l^+,\ \Psi_x = \sigma_1^3\sigma_2^3...\sigma_{l-1}^3\sigma_l^-,\ n_x = \Psi_x^+\Psi_x = \frac{1}{2}(\sigma_l^3 + 1), \qquad (5.1.7)$$

with

$$\sigma_l^\pm = \frac{1}{2}(\sigma_l^1 \pm i\sigma_l^2),\ [\sigma_l^i, \sigma_m^j] = 2i\delta_{lm}\epsilon_{ijk}\sigma_l^k. \qquad (5.1.8)$$

Here l labels the lattice point x. The Jordan-Wigner representation requires an ordering of the lattice points. For example, one can label the point $x = (x_1, x_2, x_3)$

with $x_i = 0, 1, ..., L-1$ by $l = 1 + x_1 + x_2 L + x_3 L^2$. It should be pointed out that the Jordan-Wigner representation works in any dimension. In one dimension the lattice points are, of course, naturally ordered, but even in higher dimensions the physics is completely independent of the arbitrary ordering. We now insert complete sets of fermion Fock states between the factors $\exp(-\epsilon H_i)$. Each site is either empty or occupied, i.e. n_x has eigenvalue 0 or 1. In the Pauli matrix representation this corresponds to eigenstates $|0\rangle$ and $|1\rangle$ of σ_l^3 with $\sigma_l^3|0\rangle = -|0\rangle$ and $\sigma_l^3|1\rangle = |1\rangle$. The transfer matrix is a product of factors

$$\exp(-\epsilon h_{x,i}) = \exp(\frac{\epsilon G}{4}) \begin{pmatrix} \exp(-\frac{\epsilon G}{2}) & 0 & 0 & 0 \\ 0 & \cosh\frac{\epsilon}{2} & \Sigma \sinh\frac{\epsilon}{2} & 0 \\ 0 & \Sigma \sinh\frac{\epsilon}{2} & \cosh\frac{\epsilon}{2} & 0 \\ 0 & 0 & 0 & \exp(-\frac{\epsilon G}{2}) \end{pmatrix}, \quad (5.1.9)$$

which is a 4×4 matrix in the Fock space basis $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$ of two sites x and $x + \hat{i}$. Here $\Sigma = \eta_{x,i}\sigma_{l+1}^3\sigma_{l+2}^3...\sigma_{m-1}^3$ includes the local sign $\eta_{x,i}$ as well as a non-local string of Pauli matrices running over consecutive labels between l and m, where l labels the lattice point x and m labels $x + \hat{i}$. Note that Σ is diagonal in the occupation number basis.

The partition function is now expressed as a path integral

$$Z_f = \sum_n \mathrm{Sign}[n] \exp(-S[n]), \quad (5.1.10)$$

over configurations of occupation numbers $n(x,t) = 0, 1$ on a $(3+1)$-dimensional space-time lattice of points (x,t). The Boltzmann factor

$$\begin{aligned} \exp(-S[n]) \;=\; & \prod_{\substack{x=(x_1,x_2,x_3) \\ x_1 even, t=6m}} \exp\{-s[n(x,t), n(x+\hat{1},t), n(x,t+1), n(x+\hat{1},t+1)]\} \\ & \times \prod_{\substack{x=(x_1,x_2,x_3) \\ x_2 even, t=6m+1}} \exp\{-s[n(x,t), n(x+\hat{2},t), n(x,t+1), n(x+\hat{2},t+1)]\} \\ & \times \prod_{\substack{x=(x_1,x_2,x_3) \\ x_3 even, t=6m+2}} \exp\{-s[n(x,t), n(x+\hat{3},t), n(x,t+1), n(x+\hat{3},t+1)]\} \\ & \times \prod_{\substack{x=(x_1,x_2,x_3) \\ x_1 odd, t=6m+3)}} \exp\{-s[n(x,t), n(x+\hat{1},t), n(x,t+1), n(x+\hat{1},t+1)]\} \\ & \times \prod_{\substack{x=(x_1,x_2,x_3) \\ x_2 odd, t=6m+4}} \exp\{-s[n(x,t), n(x+\hat{2},t), n(x,t+1), n(x+\hat{2},t+1)]\} \end{aligned}$$

$$\times \prod_{\substack{x=(x_1,x_2,x_3) \\ x_3 odd, t=6m+5}} \exp\{-s[n(x,t), n(x+\hat{3},t), n(x,t+1), n(x+\hat{3},t+1)]\} \tag{5.1.11}$$

is a product of space-time plaquette contributions with

$$\begin{aligned} &\exp(-s[0,0,0,0]) = \exp(-s[1,1,1,1]) = \exp(-\frac{\epsilon G}{2}), \\ &\exp(-s[0,1,0,1]) = \exp(-s[1,0,1,0]) = \cosh\frac{\epsilon}{2}, \\ &\exp(-s[0,1,1,0]) = \exp(-s[1,0,0,1]) = \sinh\frac{\epsilon}{2}. \end{aligned} \tag{5.1.12}$$

All the other Boltzmann weights are zero, which implies several constraints on allowed configurations. Note that we have dropped the trivial overall factor $\exp(\epsilon G/4)$ in eq.(5.1.9).

The sign of a configuration, Sign$[n]$, also is a product of space-time plaquette contributions $\mathrm{sign}[n(x,t), n(x+\hat{i},t), n(x,t+1), n(x+\hat{i},t+1)]$ with

$$\begin{aligned} &\mathrm{sign}[0,0,0,0]) = \mathrm{sign}[0,1,0,1]) = \mathrm{sign}[1,0,1,0]) = \mathrm{sign}[1,1,1,1]) = 1, \\ &\mathrm{sign}[0,1,1,0]) = \mathrm{sign}[1,0,0,1]) = \Sigma. \end{aligned} \tag{5.1.13}$$

It should be noted that Σ gets contributions from all lattice points with labels between l and m. This seems to make an evaluation of the fermion sign rather tedious. Also, it is not a priori obvious that Sign$[n]$ is independent of the arbitrarily chosen order of the lattice points. Fortunately, there is a simple way to compute Sign$[n]$, which is directly related to the Pauli exclusion principle and which is manifestly order-independent. In fact, Sign$[n]$ has a topological meaning. The occupied lattice sites define fermion world-lines which are closed around the Euclidean time direction. Of course, during their Euclidean time evolution fermions can interchange their positions, and the fermion world-lines define a permutation of particles. The Pauli principle dictates that the fermion sign is just the sign of that permutation. Since we work with anti-periodic spatial boundary conditions, Sign$[n]$ receives an extra minus-sign for every fermion world-line that crosses a spatial boundary.

The expectation value of a fermionic observable $O[n]$ is given by

$$\langle O \rangle_f = \frac{1}{Z_f} \sum_n O[n] \mathrm{Sign}[n] \exp(-S[n]). \tag{5.1.14}$$

Quantities of physical interest are the chiral condensate

$$\overline{\Psi}\Psi[n] = \frac{\epsilon}{6}\sum_{x,t}(-1)^{x_1+x_2+x_3}(n(x,t)-\frac{1}{2}), \tag{5.1.15}$$

and the corresponding chiral susceptibility

$$\chi = \frac{1}{\beta V}\langle(\overline{\Psi}\Psi)^2\rangle_f. \tag{5.1.16}$$

Up to now we have derived a path integral representation for the fermion system in terms of bosonic occupation numbers and a fermion sign factor that encodes Fermi statistics. The system without the sign factor is bosonic and is characterized by the positive Boltzmann factor $\exp(-S[n])$. Here the bosonic model is a quantum spin system with the Hamiltonian

$$H = \sum_{x,i}(S^1_x S^1_{x+\hat{i}} + S^2_x S^2_{x+\hat{i}} + G S^3_x S^3_{x+\hat{i}}), \tag{5.1.17}$$

where $S^i_x = \frac{1}{2}\sigma^i_l$ is a spin 1/2 operator associated with the lattice site x that was labeled by l. The case $G = 1$ corresponds to the isotropic anti-ferromagnetic quantum Heisenberg model, $G = 0$ represents the quantum XY-model, and $G = -1$ corresponds to an isotropic ferromagnet. In the language of the spin model, the chiral condensate turns into the staggered magnetization

$$\overline{\Psi}\Psi = \frac{\epsilon}{6}\sum_{x,t}(-1)^{x_1+x_2+x_3}S^3_x. \tag{5.1.18}$$

For positive G, for example for the anti-ferromagnet with $G = 1$, the staggered magnetization gets a non-zero expectation value at sufficiently low temperature, thus breaking the bosonic analog of chiral symmetry. It will turn out that the fermion sign does not change this behavior, and indeed chiral symmetry is spontaneously broken in the fermionic model as well.

5.2 The Meron-Cluster Algorithm

Let us first discuss the nature of the fermion sign problem. The fermionic path integral $Z_f = \sum_n \mathrm{Sign}[n]\exp(-S[n])$ includes the fermion sign $\mathrm{Sign}[n] = \pm 1$ as

well as a positive Boltzmann factor $\exp(-S[n])$ that contains the action $S[n]$ of the corresponding bosonic model with partition function $Z_b = \sum_n \exp(-S[n])$. A fermionic observable $O[n]$ is obtained in a simulation of the bosonic ensemble as

$$\langle O \rangle_f = \frac{1}{Z_f} \sum_n O[n] \mathrm{Sign}[n] \exp(-S[n]) = \frac{\langle O \ \mathrm{Sign} \rangle}{\langle \mathrm{Sign} \rangle}. \tag{5.2.1}$$

The average sign in the simulated bosonic ensemble is given by

$$\langle \mathrm{Sign} \rangle = \frac{1}{Z_b} \sum_n \mathrm{Sign}[n] \exp(-S[n]) = \frac{Z_f}{Z_b} = \exp(-\beta V \Delta f). \tag{5.2.2}$$

The last equality points to the heart of the sign problem. The expectation value of the sign is exponentially small in both the volume V and the inverse temperature β because the difference between the free energy densities $\Delta f = f_f - f_b$ of the fermionic and bosonic systems is necessarily positive.

Even in an ideal simulation of the bosonic ensemble which generates N completely uncorrelated configurations, the relative statistical error of the sign is

$$\frac{\Delta \mathrm{Sign}}{\langle \mathrm{Sign} \rangle} = \frac{\sqrt{\langle \mathrm{Sign}^2 \rangle - \langle \mathrm{Sign} \rangle^2}}{\sqrt{N} \langle \mathrm{Sign} \rangle} = \frac{\exp(\beta V \Delta f)}{\sqrt{N}}. \tag{5.2.3}$$

Here we have used $\mathrm{Sign}^2 = 1$. To determine the average sign with sufficient accuracy one needs to generate on the order of $N = \exp(2\beta V \Delta f)$ configurations. For large volumes and small temperatures this is impossible in practice. It is possible to solve one half of the problem if one can match any contribution -1 with another contribution 1 to give 0, such that only a few unmatched contributions 1 remain. Then effectively $\mathrm{Sign} = 0, 1$ and hence $\mathrm{Sign}^2 = \mathrm{Sign}$. This reduces the relative error to

$$\frac{\Delta \mathrm{Sign}}{\langle \mathrm{Sign} \rangle} = \frac{\sqrt{\langle \mathrm{Sign} \rangle - \langle \mathrm{Sign} \rangle^2}}{\sqrt{N'} \langle \mathrm{Sign} \rangle} = \frac{\exp(\beta V \Delta f / 2)}{\sqrt{N'}}. \tag{5.2.4}$$

One gains an exponential factor in statistics, but one still needs to generate $N' = \sqrt{N} = \exp(\beta V \Delta f)$ independent configurations in order to accurately determine the average sign. This is because one generates exponentially many vanishing contributions before one encounters a contribution 1. As explained below, in our cluster algorithm an explicit matching of contributions -1 and 1 is achieved using an improved estimator. This solves one half of the sign problem. In a second

step involving a Metropolis decision, our algorithm ensures that contributions 0 and 1 occur with similar probabilities. This saves another exponential factor in statistics and solves the other half of the sign problem.

The meron-cluster fermion algorithm is based on a cluster algorithm for the corresponding bosonic model without the sign factor. Bosonic quantum spin systems can be simulated very efficiently with the cluster algorithms. The These algorithms can be implemented directly in the Euclidean time continuum, i.e. the Suzuki-Trotter discretization is not even necessary. The same is true for the meron-cluster algorithm. Here we discuss the algorithm for discrete time. The idea behind the algorithm is to decompose a configuration into clusters which can be flipped independently. Each lattice site belongs to exactly one cluster. When the cluster is flipped, the occupation number of all the sites on the cluster is changed from $n(x,t)$ to $1-n(x,t)$, i.e. a cluster flip interchanges occupied and empty sites. The decomposition of the lattice into clusters results from connecting neighboring sites on each individual space-time interaction plaquette following probabilistic cluster rules. A sequence of connected sites defines a cluster. In this case the clusters are sets of closed loops. The cluster rules are constructed to obey detailed balance. To show this we first write the plaquette Boltzmann factors as

$$\begin{aligned}
&\exp(-s[n(x,t),n(x+\hat{i},t),n(x,t+1),n(x+\hat{i},t+1)]) = \\
&A\delta_{n(x,t),n(x,t+1)}\delta_{n(x+\hat{i},t),n(x+\hat{i},t+1)} + B\delta_{n(x,t),1-n(x+\hat{i},t)}\delta_{n(x,t+1),1-n(x+\hat{i},t+1)} + \\
&C\delta_{n(x,t),n(x,t+1)}\delta_{n(x+\hat{i},t),n(x+\hat{i},t+1)}\delta_{n(x,t),1-n(x+\hat{i},t)} \\
&\quad +D\delta_{n(x,t),n(x+\hat{i},t+1)}\delta_{n(x+\hat{i},t),n(x,t+1)} + \\
&E\delta_{n(x,t),n(x,t+1)}\delta_{n(x+\hat{i},t),n(x+\hat{i},t+1)}\delta_{n(x,t),n(x+\hat{i},t)}. \qquad (5.2.5)
\end{aligned}$$

The various δ-functions specify which sites are connected and thus belong to the same cluster. The quantities $A, B, ..., E$ determine the relative probabilities for different cluster break-ups of an interaction plaquette. We only allow break-ups which generate legal configurations under cluster flips. For example, A determines the probability with which sites are connected with their time-like neighbors, while B and D determine the probabilities for connections with space-like or diagonal neighbors, respectively. The quantities C and E determine the probabilities to put all four sites of a plaquette into the same cluster. This is possible for plaquette configurations $[0,1,0,1]$ or $[1,0,1,0]$ with a probability proportional to C and for configurations $[0,0,0,0]$ or $[1,1,1,1]$ with a probability proportional to E.

Inserting the expressions from eq.(5.1.12) one finds

$$\begin{aligned}
&\exp(-s[0,0,0,0]) = \exp(-s[1,1,1,1]) = \exp(-\frac{\epsilon G}{2}) = A + D + E, \\
&\exp(-s[0,1,0,1]) = \exp(-s[1,0,1,0]) = \cosh\frac{\epsilon}{2} = A + B + C, \\
&\exp(-s[0,1,1,0]) = \exp(-s[1,0,0,1]) = \sinh\frac{\epsilon}{2} = B + D. \qquad (5.2.6)
\end{aligned}$$

For example, the probability to connect the sites with their time-like neighbors on a plaquette with configuration $[0,0,0,0]$ or $[1,1,1,1]$ is $A/(A+D+E)$, while the probability for a connection with their diagonal neighbor is $D/(A+D+E)$. All sites on such a plaquette are put into the same cluster with probability $E/(A+D+E)$. Similarly, the probability for connecting space-like neighbors on a plaquette with configuration $[0,1,1,0]$ or $[1,0,0,1]$ is $B/(B+D)$ and the probability for diagonal connections is $D/(B+D)$.

Eq.(5.2.5) can be viewed as a representation of the original model as a random cluster model. The cluster algorithm operates in two steps. First, a cluster break-up is chosen for each space-time interaction plaquette according to the above probabilities. This effectively replaces the original Boltzmann weight of a plaquette configuration with a set of constraints represented by the δ-functions associated with the chosen break-up. The constraints imply that occupation numbers of connected sites can only be changed together. In the second step of the algorithm every cluster is flipped with probability $1/2$. When a cluster is flipped the occupation numbers of all sites that belong to the cluster are changed. Eq.(5.2.6) ensures that the cluster algorithm obeys detailed balance. To determine $A, B, ..., E$ we distinguish three cases. For $G \geq 1$ we solve eq.(5.2.6) by

$$A = \exp(-\frac{\epsilon G}{2}),\ B = \sinh\frac{\epsilon}{2},\ C = \exp(-\frac{\epsilon}{2}) - \exp(-\frac{\epsilon G}{2}),\ D = E = 0. \quad (5.2.7)$$

For $-1 \leq G \leq 1$ we use

$$\begin{aligned}
&A = \frac{1}{2}[\exp(-\frac{\epsilon G}{2}) + \exp(-\frac{\epsilon}{2})],\ B = \frac{1}{2}[\exp(\frac{\epsilon}{2}) - \exp(-\frac{\epsilon G}{2})],\ C = 0, \\
&D = \frac{1}{2}[\exp(-\frac{\epsilon G}{2}) - \exp(-\frac{\epsilon}{2})],\ E = 0, \qquad (5.2.8)
\end{aligned}$$

and, finally, for $G \leq -1$

$$A = \cosh\frac{\epsilon}{2},\ B = C = 0,\ D = \sinh\frac{\epsilon}{2},\ E = \exp(-\frac{\epsilon G}{2}) - \exp(\frac{\epsilon}{2}). \qquad (5.2.9)$$

As an example, let us consider the anti-ferromagnetic quantum Heisenberg model, i.e. $G = 1$, and hence

$$A = \exp(-\frac{\epsilon}{2}),\ B = \sinh\frac{\epsilon}{2},\ C = D = E = 0. \tag{5.2.10}$$

Thus, on plaquette configurations $[0,0,0,0]$ or $[1,1,1,1]$ one always chooses time-like connections between sites, and for configurations $[0,1,1,0]$ or $[1,0,0,1]$ one always chooses space-like connections. For configurations $[0,1,0,1]$ or $[1,0,1,0]$ one chooses time-like connections with probability $p = A/(A+B) = 2/[1+\exp(G/2)]$ and space-like connections with probability $1-p = B/(A+B)$. This algorithm is extremely efficient and has almost no detectable auto-correlations, and its dynamical exponent for critical slowing down is compatible with zero, i.e. $z \approx 0$. Later it will be important that diagonal connections are never chosen for $G \geq 1$.

Let us now consider the effect of a cluster flip on the fermion sign. It is obvious that the flip of a cluster either leads to a sign change or it leaves the sign unchanged. In general, the effect of the flip of a specific cluster on the fermion sign depends on the orientation of the other clusters. For example, a cluster whose flip does not change the sign now, may very well change the sign after other clusters have been flipped. In other words, the clusters affect each other in their effect on the fermion sign. This makes it very difficult to understand the effect of the various cluster flips on the topology of the fermion world-lines and thus on Sign$[n]$. As a consequence, for the most general model described above, we don't know how to solve the fermion sign problem. However, there are clusters whose effect on Sign$[n]$ is independent of the orientation of all the other clusters. Specifically, these are those clusters that do not contain any diagonal break-ups. To ensure that no clusters contain such break-ups we limit ourselves to models which have $D = 0$. According to eq.(5.2.7) this is the case for $G \geq 1$.

Once we forbid diagonal cluster break-ups, i.e. when $D = 0$, the clusters have a remarkable property with far reaching consequences: each cluster can then be characterized by its effect on the fermion sign independent of the orientation of all the other clusters. We refer to clusters whose flip changes Sign$[n]$ as merons, while clusters whose flip leaves Sign$[n]$ unchanged are called non-merons. The flip of a meron-cluster permutes the fermions and changes the topology of the fermion world-lines. The term meron has been used before to denote half-instantons. For example, the meron-clusters in the algorithm for the 2-d $O(3)$ model at non-zero vacuum angle θ are indeed half-instantons. The number of merons in a

configuration is always even. This follows when one considers flipping all clusters, and thus changing the occupation of all lattice sites, which leaves the fermion sign unchanged. If the number of merons would be odd, flipping all clusters would change the fermion sign, which implies that the number of merons must be even.

The meron concept alone allows us to gain an exponential factor in statistics. Since all clusters can be flipped independently, one can construct an improved estimator for Sign$[n]$ by averaging analytically over the 2^{N_C} configurations obtained by flipping the N_C clusters in the configuration in all possible ways. For configurations that contain merons the average Sign$[n]$ is zero because flipping a single meron leads to a cancellation of contributions ± 1. Hence, only the configurations without merons contribute to Sign$[n]$. The vast majority of configurations contains merons and now contributes an exact 0 to Sign$[n]$ instead of a statistical average of contributions ± 1. In this way the improved estimator leads to an exponential gain in statistics.

Still, as it stands it is not guaranteed that the contributions from the zero-meron sector will always be positive. With no merons in the configuration it is clear that the fermion sign remains unchanged under cluster flip, but one could still have Sign$[n] = -1$. Fortunately, there is a way to guarantee that Sign$[n] = 1$ in the zero-meron sector. In that case the contributions to Sign$[n]$ are 0 from configurations containing meron-clusters and 1 from configurations without merons. According to the previous discussion, this solves one half of the fermion sign problem. Let us describe how one can make sure that a configuration without merons always has Sign$[n] = 1$. This is possible when one does not allow the cluster break-ups characterized by the amplitudes D and E in eq.(5.2.5). This again limits us to models with $G \geq 1$, for which we have $D = E = 0$ according to eq.(5.2.7). The remaining cluster break-ups with amplitudes A, B, C have a very important property. They guarantee that sites inside a cluster obey a pattern of staggered occupation, i.e. either the even sites (with $x_1 + x_2 + x_3$ even) within the cluster are all occupied and the odd sites are all empty, or the even sites are all empty and the odd sites are all occupied. This guarantees that the clusters can be flipped such that one reaches the totally staggered configuration in which at all times all even sites are occupied and all odd sites are empty. In this half-occupied configuration all fermions are static, no fermions are permuted during the Euclidean time evolution, and thus Sign$[n] = 1$. Since any configuration can be turned into the totally staggered configuration by appropriate cluster flips, this is particularly true for configurations without merons. Since the totally staggered configuration has Sign$[n] = 1$ and the fermion sign remains unchanged when a

non-meron-cluster is flipped, all configurations without merons have $\mathrm{Sign}[n] = 1$. Instead of a sequence of ± 1 for $\mathrm{Sign}[n]$, we now have contributions 0 and 1. As discussed before, this only solves one half of the fermion sign problem. Before we can solve the other half of the problem we must discuss improved estimators for the physical observables.

Let us consider an improved estimator for $(\overline{\Psi}\Psi[n])^2\mathrm{Sign}[n]$ which is needed to determine the chiral susceptibility χ. The total chiral condensate, $\overline{\Psi}\Psi[n] = \sum_C \overline{\Psi}\Psi_C$, is a sum of cluster contributions, $\overline{\Psi}\Psi_C = \epsilon \sum_{(x,t)\in C}(-1)^{x_1+x_2+x_3}(n(x,t)-\frac{1}{2})$. When a cluster is flipped, its condensate contribution changes sign. In a configuration without merons, where $\mathrm{Sign}[n] = 1$ for all relative cluster flips, the average of $(\overline{\Psi}\Psi[n])^2\mathrm{Sign}[n]$ over all 2^{N_C} configurations is $\sum_C |\overline{\Psi}\Psi_C|^2$. For configurations with two merons the average is $2|\overline{\Psi}\Psi_{C_1}||\overline{\Psi}\Psi_{C_2}|$ where $C_1$and C_2 are the two meron-clusters. Configurations with more than two merons do not contribute to $(\overline{\Psi}\Psi[n])^2\mathrm{Sign}[n]$. The improved estimator for the susceptibility is hence given by

$$\chi = \frac{\langle \sum_C |\overline{\Psi}\Psi_C|^2 \delta_{N,0} + 2|\overline{\Psi}\Psi_{C_1}||\overline{\Psi}\Psi_{C_2}|\delta_{N,2}\rangle}{V\beta\langle\delta_{N,0}\rangle}, \tag{5.2.11}$$

where N is the number of meron-clusters in a configuration. Thus, to determine χ one must sample the zero and two-meron sectors.

The probability to find a configuration without merons is exponentially small in the space-time volume since it is equal to $\langle\mathrm{Sign}\rangle$. Thus, although we have increased the statistics tremendously with the improved estimators, without a second step one would still need an exponentially large statistics to accurately determine χ. Fortunately, the numerator in equation (5.2.11) receives contributions from the zero- and two-meron sectors only, while the denominator gets contributions only from the zero-meron sector. One can hence restrict oneself to the zero- and two-meron sectors and never generate configurations with more than two merons. This enhances both the numerator and the denominator by a factor that is exponentially large in the volume, but leaves the ratio of the two invariant. One purpose of the second step of the meron cluster algorithm is to eliminate all configurations with more than two merons. To achieve this, we start with an initial configuration with zero or two merons. For example, a completely occupied configuration has no merons. We then visit all plaquette interactions one after the other and choose new pair connections between the four sites according to the above cluster rules. If the new connection increases the number of merons beyond two, it is not accepted and the old connection is kept for that

plaquette. This procedure obeys detailed balance because configurations with more than two merons do not contribute to the observable we consider. This simple reject step eliminates almost all configurations with weight 0 and is the essential step to solve the other half of the fermion sign problem.

Assuming a dilute gas of meron and non-meron-clusters of typical space-time volume $|C|$ one finds a ratio $p(0)/p(2) \propto (|C|/V\beta)^2$ of the probabilities $p(0)$ and $p(2)$ to have zero or two merons. Hence, as long as the cluster size does not grow with the space-time volume, most configurations would have two merons and therefore would still have weight 0. Without further improvements, one would still need statistics quadratic (but no longer exponential) in the space-time volume to get an accurate average sign. The remaining problem can be solved with a re-weighting technique. To enhance the zero-meron configurations in a controlled way, we introduce trial probabilities $p_t(0)$ and $p_t(2)$ which determine the relative weight of the zero and two-meron sector. The trial distribution $p_t(N)$ for $N > 2$ is set to infinity. The distribution $p_t(N)$ is used in a Metropolis accept-reject step for the newly proposed pair connection on a specific plaquette interaction. A new pair connection that changes the meron number from N to N' is accepted with probability $p = \min[1, p_t(N)/p_t(N')]$. In particular, configurations with $N' > 2$ are never generated because then $p_t(N') = \infty$ and $p = 0$. After visiting all plaquette interactions, each cluster is flipped with probability $1/2$ which completes one update sweep. After re-weighting, the zero and two-meron configurations appear with similar probabilities. This completes the second step in our solution of the fermion sign problem. The re-weighting of the zero and two-meron configurations is taken into account in the final expression for the susceptibility as

$$\chi = \frac{\langle \sum_C |\overline{\Psi}\Psi_C|^2 \delta_{N,0}\ p_t(0) + 2|\overline{\Psi}\Psi_{C_1}||\overline{\Psi}\Psi_{C_2}|\delta_{N,2}\ p_t(2)\rangle}{V\beta\langle \delta_{N,0}\ p_t(0)\rangle}. \tag{5.2.12}$$

The optimal ratio $p_t(0)/p_t(2)$, which minimizes the statistical error of χ, can be estimated by gradually increasing the volume and the inverse temperature to their desired values.

5.3 Numerical Results

Figure 5.1 shows the chiral susceptibility χ as a function of β for various spatial sizes L. At high temperatures (small β) χ is almost independent of the volume,

indicating that chiral symmetry is unbroken. On the other hand, at low temperatures *chi* increases with the volume, which implies that chiral symmetry is spontaneously broken.

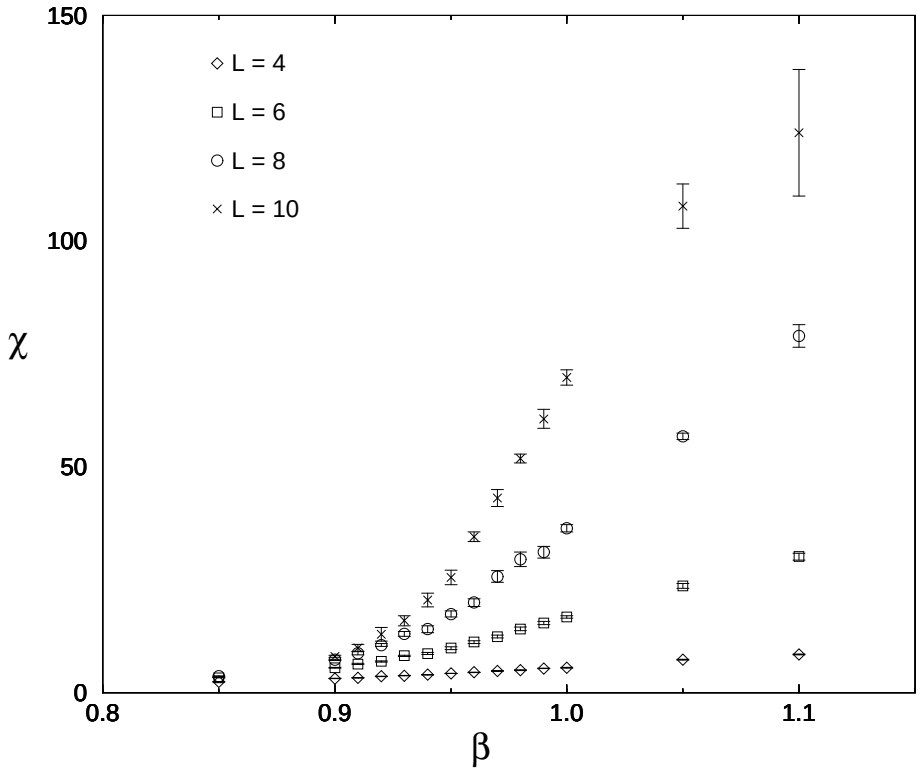

Figure 5.1: The chiral susceptibility χ as a function of the inverse temperature β for various spatial sizes $L = 4, 6, 8, 10$.

To study the critical behavior in detail, we have performed a finite-size scaling analysis for χ focusing on a narrow range $\beta \in [0.9, 1.0]$ around the critical point, which turns out to be at $\beta_c = 0.948(2)$. Since a $\mathbf{Z}(2)$ chiral symmetry gets spontaneously broken at a finite temperature in this $(3+1)$-dimensional model, one expects to find the critical behavior of the 3-d Ising model. The corresponding finite-size scaling formula is

$$\chi(L,\beta) = a(\beta) + b((\beta - \beta_c)L^{1/\nu})L^{\gamma/\nu},$$
$$a(\beta) = a_0 + a_1(\beta - \beta_c) + ...,$$

$$b(\beta) = b_0 + b_1(\beta - \beta_c)L^{1/\nu} + b_2(\beta - \beta_c)^2 L^{2/\nu} + ... \qquad (5.3.1)$$

For the 3-d Ising model the critical exponents are given by $\nu = 0.630(1)$ and $\gamma/\nu = 1.963(3)$. Fitting our data, we find $\nu = 0.69(2)$ and $\gamma/\nu = 1.98(2)$, which indicates that the chiral transition of the staggered fermion model is indeed in the 3-d Ising universality class. The fit gives $\beta_c = 0.948(3)$. In figure 5.2 we have taken the values of the critical exponents from the 3-d Ising model, and we have plotted $\chi/L^{\gamma/\nu}$ as a function of $(\beta - \beta_c)L^{1/\nu}$. We have varied the value of β_c and found that indeed all data can be collapsed on one universal curve. The resulting critical inverse temperature is $\beta_c = 0.948(3)$ in agreement with the previous fit. At the estimated value of β_c, we have performed simulations on larger spatial

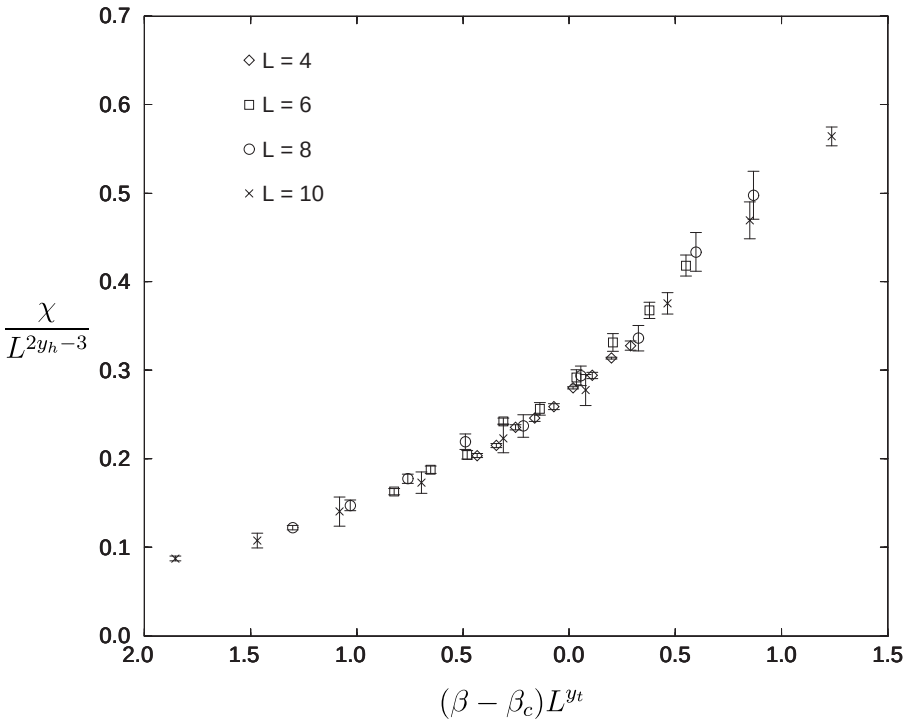

Figure 5.2: Finite-size scaling behavior of the chiral susceptibility χ. Data for various spatial sizes $L = 4, 6, 8, 10$ fall on one universal curve.

volumes up to 16^3. The results for χ are shown in figure 5.3 together with a fit

that gives an independent estimate of $\gamma/\nu = 1.98(1)$. All of this supports the claim that the chiral transition in our model is Ising-like.

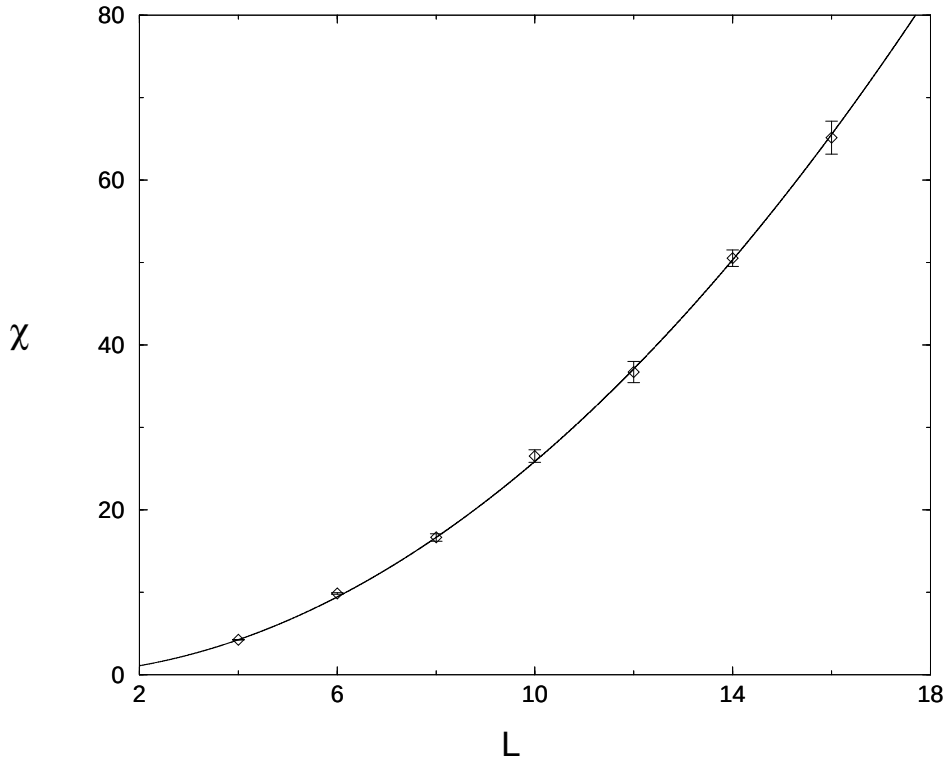

Figure 5.3: Finite-size scaling behavior of the chiral susceptibility χ as a function of the spatial size L at the critical inverse temperature β_c. A fit of the volume-dependence gives the critical exponent $\gamma/\nu = 1.98(1)$ of the 3-d Ising model.

It should be pointed out that the staggered fermion model suffers from a very severe sign problem. For example, an extrapolation from smaller volumes gives the rough estimate $\langle \mathrm{Sign} \rangle \approx 10^{-50}$ for the 16^3 lattice. Hence, to achieve a similar accuracy without the meron-cluster algorithm one would have to increase the statistics by a factor 10^{100}, which is obviously impossible. In fact, at present there is no other method that can simulate this model.

5.4 Conclusions

We have applied a new fermion simulation technique — the meron-cluster algorithm — to a model of staggered fermions. In contrast to standard methods which integrate out the fermions and are left with a non-local bosonic action, we describe the fermions in a Fock state occupation number basis and thus keep a local bosonic action. Still, the fermion permutation sign arises due the Pauli exclusion principle as a non-local factor associated with non-trivial topology of the fermion world-lines. The sign results in severe cancellations which makes standard simulation techniques impossible to use. The decomposition of a configuration into clusters allows us to disentangle the complicated topology of the fermion world-lines. In particular, flipping a meron-cluster leads to a permutation of fermions and thus to an explicit cancellation of two contributions ± 1 to the path integral, such that only the zero-meron sector contributes to the partition function. Observables like the chiral susceptibility χ receive contributions only from the zero- and two-meron sectors. To measure χ one can hence eliminate all sectors with more than two merons, which leads to an exponential gain in statistics and to a complete solution of the fermion sign problem.

The meron-cluster algorithm allowed us to simulate a staggered fermion model which suffers from a severe sign problem. The model has two flavors and a $\mathbb{Z}(2)$ chiral symmetry which is spontaneously broken at low temperatures. Applying finite-size scaling methods to the high-precision numerical data for the chiral susceptibility, we extracted critical exponents compatible with those of the 3-d Ising model. This is the expected behavior based on universality arguments. It would be interesting to apply our method to (2+1) dimensions. In the $N = 4$ flavor case was studied it was verified that the model is in the 2-d Ising universality class. The standard fermion algorithm that was used in that study does not work for $N < 4$ due to the fermion sign problem. On the other hand, the meron-cluster algorithm solves the fermion sign problem and can be used to explore those models.

Even in cases without a sign problem, the meron-cluster algorithm is more efficient than standard fermion simulation methods. Unfortunately, the meron-cluster algorithm is not universally applicable. For example, the meron concept applies only when the clusters are independent in their effect on the fermion sign. In addition, it must always be possible to flip the clusters such that one reaches a reference configuration with a positive sign. Otherwise, some contributions

from the zero-meron sector could be negative. For example, in our model these restrictions led to $G \geq 1$, i.e. to a sufficiently strong four-fermion coupling. We have not attempted to take the continuum limit of the lattice theory. Instead, we have studied the chiral phase transition at a finite lattice spacing corresponding to $G = 1$. The restriction $G \geq 1$ would prevents us from approaching the continuum limit if it corresponds to $G_c < 1$. It should be pointed out that we can study the universal behavior of the chiral transition without taking the continuum limit. Of course, it is not excluded that appropriate modifications of the algorithm might allow us to work at $G < 1$.

Chapter 6

Exercises

The purpose of these exercises is to learn the Monte Carlo method using the Ising model as a simple physical system of classical statistical mechanics. To achieve our goal it is indispensable that one develops and debugs Monte Carlo routines on his or her own. Here are some exercises that describe a useful way to proceed.

6.1 Metropolis Algorithm for the 1-d Ising Model

The purpose of the first exercise is to develop a Metropolis Monte Carlo code for the 1-d Ising model. It would be useful to have a MAIN program from which appropriate subroutines are called. In the MAIN program one could define a field (e.g. SPIN(N)) that contains the spin value for each lattice point N. It is also useful to define fields that contain the left and right neighbors of a given lattice point (e.g. ILEFT(N) and IRIGHT(N)). In this way periodic boundary conditions are easily implemented.

The MAIN program should call a subroutine START which sets an initial spin configuration. In principle, it does not matter which initial configuration one picks. However, it will be a good check of the procedure that the final answers should be statistically independent of the initial configuration. Hence, it would be useful to allow for different initial configurations, e.g. a completely ordered one with all spins parallel, or a random one. In order to assign a random

value to an Ising spin, one can put it to 1 if a random number is larger than 1/2 and to -1 otherwise.

The program should then contain a subroutine METROPOLIS in which one Metropolis sweep through the lattice is performed. In one sweep each spin is subjected to one Metropolis update, i.e. its neighbors are examined and the change of the energy under spin flip is calculated. Then the spin is flipped or not, according to the rules of the Metropolis algorithm. If the spin should be flipped with probability p, one can pick a random number x and flip the spin if $x < p$.

There could also be a subroutine MAGNETIZATION which measures the magnetization of a given configuration. This value is then used in the computation of the susceptibility. An even more efficient method is to keep track of the magnetization after each individual spin update.

An important part of each Monte Carlo routine is the error analysis. When the Monte Carlo data are written into a file, the error analysis can be performed with a separate program.

Finally, we like to compare the results of our Monte Carlo simulation with the exact analytic results. For this purpose we need a separate routine that evaluates the exact result numerically. One can then compare Monte Carlo and analytic results and decide if they agree within the statistical errors. At first they will usually not and one must start the sometimes painful process of debugging the code.

6.2 Metropolis Algorithm for the 2-d Ising Model

The Monte Carlo program for the 2-d Ising model can be based on the one for the 1-d model and requires only minor modifications. In particular, the neighbor lists must be modified and the change of the energy under spin flip must take into account the interactions of a spin with all four neighbors.

One can now measure the susceptibility and show that it peaks near the analytically known critical temperature. It is also useful to monitor the statistical errors as one approaches the critical temperature. One will then see that the Metropolis algorithm becomes rather inefficient. It suffers from critical slowing

down, i.e. the number of sweeps between statistically independent spin configurations increases dramatically when one approaches criticality. This inefficiency of the Metropolis algorithm near the critical point is a good motivation for moving on to the much more efficient Swendsen-Wang and Wolff cluster algorithms.

6.3 Swendsen-Wang Cluster Algorithm for the Ising Model

The next exercise is to construct a program for the Swendsen-Wang cluster algorithm both for the 1-d and the 2-d Ising model. The Swendsen-Wang algorithm is a multi-cluster algorithm, i.e. all clusters in a configuration are identified and then flipped with 50 percent probability. One can use certain elements of the Metropolis code. For example, the spin configuration and the neighbor relations can be stored in the same way. In addition, we now also need to store the bond configuration. Also one may want to introduce a mark for each spin which denotes the number of the cluster to which it belongs. The start configuration can be generated as before.

In the Swendsen-Wang algorithm one first goes through all bonds sequentially. Based on the spins at the two ends of a bond, one decides if the bond is activated or not. Antiparallel spins are never connected by an activated bond, while parallel spins are connected with the appropriate probability. After all bonds have been activated or deactivated, one identifies the clusters. For this purpose one starts with a given spin site and puts it as the first member of a cluster list. One also marks this site with the number of the corresponding cluster. Then one examines all the bonds connected to the selected site. If a bond is activated, the corresponding neighboring spin site is included in the list of cluster members. Also the corresponding site is marked with the number of the corresponding cluster. In this way one can ensure that a site is not put on the cluster list more than once. One now moves on to the next member on the list and examines its neighbors until all members of the cluster have been identified. Then the cluster size is recorded in order to construct the improved estimator for the susceptibility, and all spins on the cluster are flipped with 50 percent probability. Once all clusters in the configuration have been identified and have eventually been flipped, a new bond configuration is generated and the procedure is repeated.

The cluster algorithm can again be checked against the exact results in one dimension. Once the code is debugged, it can be applied to the 2-d Ising model near criticality. Again one can determine the autocorrelation time as one approaches the critical temperature. It will then become clear that the cluster algorithm is much more efficient than the Metropolis algorithm.

6.4 Wolff Single-Cluster Algorithm for the Ising Model

The Wolff cluster algorithm is a single-cluster algorithm, i.e. only a single cluster is constructed for each configuration, but it is flipped with 100 percent probability. In this case, it would be inefficient to activate or deactivate all bonds in the configuration. Instead one decides if a bond is activated only when one asks if the corresponding neighbor is a member of the cluster. One still must remember if a bond has already been activated or not, in case this bond is visited a second time. Once all spins of the cluster have been identified, they are flipped with 100 percent probability. Alternatively, the spins can already be flipped immediately after they have been identified as cluster members. Again, the cluster size is recorded in order to identify the improved estimator for the susceptibility.

It should be noted that the concept of a sweep is a bit subtle for the Wolff cluster algorithm. In that case, depending on the cluster size, one should define a sweep as the number of cluster flips after which on average the total number of spins in the lattice has been flipped once.

6.5 Swendsen-Wang Cluster Algorithm for the Potts Model

In this exercise, a multi-cluster algorithm is constructed for the Potts model. Instead of the two states for the Ising model, a Potts spin can exist in N different states. As before, all spins in a cluster have the same value and are flipped together. In this case, one of the other $N-1$ values is selected at random. The probability for putting a bond between like spins is again determined by the

requirement of detailed balance. It is straightforward to operate the algorithm also in the Wolff single-cluster version.

6.6 Severity of the Complex Action Problem

In order to appreciate the exponential severity of the complex action problem in the Potts model for dense QCD, this exercise consists of the measurement of the complex part of the Boltzmann factor as an observable. In particular, one should convince oneself that the signal degrades exponentially compared to the noise as one increases the volume. This study can be performed, for example, with the multi-cluster algorithm. One may also attempt to measure the expectation value of the Polyakov loop for non-zero chemical potential μ. In practice this will be possible only for small μ or for small volume because the signal to noise ratio is again exponentially small.

6.7 Cluster Algorithm for the Potts Model at Non-zero μ

In the final exercise the chemical potential shall be included in the Potts model. The multi-cluster algorithm is then turned into a meron-cluster algorithm which is based on the improved estimator for the complex part of the Boltzmann weight. Each configuration then has a weight that depends on the cluster sizes. The cluster bonds are updated sequentially respecting those weights. In particular, one must ask if it is acceptable to glue two different cluster together or to separate one cluster into two pieces. This decision is made in a Metropolis step. At the end, one may now convince oneself that the complex action problem is indeed solved in this way.