Asymptotic Scaling in the Two-Dimensional O(3) Nonlinear Sigma Model: A Monte Carlo Study On Parallel Computers

Thesis by
John Apostolakis
In Partial Fulfillment of the Requirements
for the Degree of Doctor
of Philosophy

California Institute of Technology
Pasadena, California

1994
(Submitted August 11, 1993)

Acknowledgements

It is a great pleasure and at the same time a very hard task to thank, appropriately, all those who helped me in this quest. First I must thank my parents for their love and constant support all these 28 years and for giving me so much that another 120 pages couldn't really properly describe it. I feel I owe great thanks to my advisor, Professor Geoffrey C. Fox, who provided me the chance and the means to accomplish this work, and even more importantly the encouragement and steering needed to do it. I would also especially like to thank Professors John Preskill, Steve Frautchi, and Enzo Marinari, who provided me with help, understanding, advice and impetus during my final years as a graduate student.

In my long time at Caltech and at Syracuse there were so many friends with whom I shared my experiences that it would be an almost impossible task to try to name them all. But since I feel I must try to remember some of you, here is my attempt. If it were not for you, my sanity (whatever I have), would certainly have not survived. Thanks Elias Kiritsis, Alexis Polychronakos, Giorgos Siopsis, Paul Coppi, Chris S. Kochanek, Sandip Trivedi, Ted Allen, Ian Angus, Brian Warr, Michele Sere, Chris Wilson, Alain Picard, Rich Rand, Julie Moses, Bob Wilson, Doug Collins, Tom Frey, Helen Johnston, Chris Tinney, Neil Reidd, Mark Muldoon, JoAnn Boyd, Stuart Anderson, Darin Beigie, Paul Ray, Aris Moustakas, Panos Pantzikas, Petros Mouhtaris, Nikos Mpekiaris, and all the innumerable people with whom I played softball, soccer, volleyball and basketball at Tech.

After my journey east there were many who helped make life bearable at first and then live-able again. My neighbors and very good friends Homer Antoniades, Elda Tzoumaka and Ted, the many members of the Syracuse University Outing Club and in particular the great caving leaders, Kostas Anagnostopoulos, Nikos Kalogeropoulos, Nikos Troulinos, Antonis Patrinos, Kostas and all the great soccer gang at Vincent, Marco Falcioni, Geoff Harris and Zheng-Yao Su.

In addition there are several people with whom, in addition to sharing friendship, I had the great chance to work with. For all you, many thanks: Jon Flower, C.S.K., Clive Baillie, Miloje Makivic, H.Q. Ding, Paul Coddington and Enzo M. Also greatly deserved thanks to my teachers at Tech - K. Thorne, D. Hitlin, J. Preskill, P. Goldreich, R. Blanford, F. Porter, F. Zachariasen, M. Cross and M. Wise.

Few could have been so lucky as to have known as great and encouraging people as Donna Driscoll and Deborah Jones in their graduate career. Of the several others who's friendly help in the more practical aspects of University life was given in large measure, I would particularly like to thank Diane Sanderson, and, for help beyond the call of computing support duty, Mark Levinson and Doug Freyberger.

To the many others I should also acknowledge, please know I remember you when my memory is not as fogged up as it is right now, and thanks!

Finally, I would particularly like to thank my warm and loving companion,

Sharon, who has made the past year the happiest I could imagine, even though it has been a time of ever increasing academic and job seeking stress.

Abstract

We investigate the 2d O(3) model with the standard action by Monte Carlo simulation at couplings β up to 2.05 . We measure the energy density, mass gap and susceptibility of the model, and gather high statistics on lattices of size $L \leq 1024$ using the Floating Point Systems T-series vector hypercube and the Thinking Machines Corp.'s Connection Machine 2. Asymptotic scaling does not appear to set in for this action, even at $\beta = 2.10$, where the correlation length is 420. We observe a 20% difference between our estimate $m/\Lambda_{\overline{MS}} = 3.52(6)$ at this β and the recent exact analytical result. We use the overrelaxation algorithm interleaved with Metropolis updates and show that decorrelation time scales with the correlation length and the number of overrelaxation steps per sweep. We determine its effective dynamical critical exponent to be $z' = 1.079(10)$; thus critical slowing down is reduced significantly for this local algorithm that is vectorizable and parallelizable.

We also use the cluster Monte Carlo algorithms, which are non-local Monte Carlo update schemes which can greatly increase the efficiency of computer simulations of spin models. The major computational task in these algorithms is connected component labeling, to identify clusters of connected sites on a lattice. We have devised some new SIMD component labeling algorithms, and implemented them on the Connection Machine. We investigate

their performance when applied to the cluster update of the two dimensional Ising spin model.

Finally we use a Monte Carlo Renormalization Group method to directly measure the couplings of block Hamiltonians at different blocking levels. For the usual averaging block transformation we confirm the renormalized trajectory (RT) observed by Okawa. For another improved probabilistic block transformation we find the RT, showing that it is much closer to the Standard Action. We then use this block transformation to obtain the discrete β-function of the model which we compare to the perturbative result. We do not see convergence, except when using a rescaled coupling β_E to effectively resum the series. For the latter case we see agreement for $m/\Lambda_{\overline{MS}}$ at $\beta = 2.14, 2.26, 2.38$ and 2.50. To three loops $m/\Lambda_{\overline{MS}} = 3.047(35)$ at $\beta = 2.50$, which is very close to the exact value $m/\Lambda_{\overline{MS}} = 2.943$. Our last point at $\beta = 2.62$ disagrees with this estimate however.

Contents

Chapter 1

Introduction

Our best current theory of nature, the Standard Model, is a gauge theory. Four dimensional gauge theories thus form an essential part of our understanding of natural phenomena, but they are complicated and our comprehension of them is lacking in some respects. In particular non-Abelian gauge theories can provide analytical predictions only at very high energies where the coupling constant becomes small. This important phenomenon is called asymptotic freedom, and plays a key part in our ability to compare between experiment and theory.

At lower energies analytical calculations can only be carried out in certain approximations and for simple models like, e.g., the non relativistic quark model. Thus important properties, like dynamical mass generation, and the

resulting physical quantities like, e.g., the ratio of the masses of nucleons, which stem from the infrared behavior of the theory can only be computed from first principles by using lattice gauge theory.

There is another way to study these properties though, which is to study models that exhibit many similar properties. In fact, part of our knowledge of four-dimensional gauge theories is derived from the study of these simpler models. One such class of models, the nonlinear sigma models in two dimensions have been studied extensively.

Two-dimensional non linear sigma models, in common with four-dimensional gauge theories, are asymptotically free in the ultraviolet (high energy limit), while in the infrared (at large distances) they are strongly coupled and generate a mass gap dynamically; some also have instanton solutions. Studying the simpler models lends important insight into the study of the more complicated gauge theories. For this reason they have also been used to develop and evaluate many new methods later used to study gauge theories.

The $O(N)$ non linear sigma models are a class of two-dimensional models that, for $N > 2$, are asymptotically free. They include the simplest model that possesses that property, the $O(3)$ model. The $O(3)$ model, but not the others, also possesses instanton solution in common with four-dimensional

gauge theories.

Two-dimensional $O(N)$ models, are related to lattice models known as n-vector models or $O(N)$ invariant classical Heisenberg models. Our understanding of both of these models in the weak coupling regime (low temperature in the classical Heisenberg model) originates in the Renormalization Group treatment of Polyakov[1]. To study the long wavelength limit he integrated out the high momentum modes and showed that the beta function is negative at the weakest couplings. Thus by starting at weak coupling and investigating the behavior at larger length scales the running coupling should be driven into the regime of strong coupling (high temperature), which is known to have exponential correlation functions[2]. The crossover between weak and strong coupling was observed and the prediction of a single phase was confirmed by the pioneering simulation of Shenker and Tobochnik[3] using Monte Carlo and Monte Carlo Renormalization Group methods.

Subsequent Renormalization Group calculations determined the β function and the anomalous dimension of the field γ to two-loops [4], the extent of its universal part. From these the behavior of the mass gap and susceptibility of the model in two dimensions can be obtained. The mass gap m is a multiple of the dynamically generated mass scale Λ : $m = c_A \Lambda_A$, where the subscript A is intended to show which quantities depend on the regularization

(or lattice action) chosen. Taking into account the subsequent calculations to three loops (made first in the modified minimal subtraction scheme[5] and later for some lattice actions[6]) the scale is

$$\Lambda_A(\beta_A) = 2\pi\beta_A e^{-2\pi\beta_A}\left(1+\frac{\delta_A}{2\pi\beta_A}\right). \tag{1.1}$$

Thus, for different regularizations, the scale Λ, the value of δ and the ratio m/Λ change; however the different values of Λ can been related in a perturbative calculation. This was done first by Parisi[7], who related the parameter for the standard, nearest neighbor, lattice action Λ_{SA} and that for the modified minimal subtraction regularization of the continuum model $m/\Lambda_{\overline{MS}}$, yielding $m/\Lambda_{\overline{MS}}/\Lambda_{SA} = 27.21$.

Later Monte Carlo simulations attempted to find scaling of the mass gap and susceptibility of the model according to the two-loop β function[8, 9, 10, 11, 12], scaling which is known as asymptotic scaling. Using the nearest neighbor action they also confirmed the qualitative nature of the predictions but either did not see asymptotic scaling[8, 9, 10, 12], or observed convergence to asymptotic scaling for the mass gap[11] (for $\beta \geq 1.6$) within large statistical errors. The value obtained for the proportionality constant $m/\Lambda_{SA} = 110 \pm 10$ was at a variance with that obtained by other methods,

notably large N and small volume expansions. Also recent high precision Monte Carlo calculations, particularly that of Wolff[13] and ours[14] (see chapter 4), have shown that more precise estimates show large deviations from asymptotic scaling in this region.

Several other methods have been used to probe the properties of the model. The original high temperature expansions and the strong coupling expansion demonstrated the existence of exponential correlations in that temperature region.

Our knowledge of the behavior of the model has also been greatly enhanced by the investigation of other aspects of the model. It has been shown to have an infinite set of non-local conservation laws. These in turn lead to the lack of particle production and conservation of particle number in particle interactions and factorizability of the scattering matrix. Using this property, Zamolodchikov[15] proposed an $O(N)$ invariant two particle S matrix and showed it was the correct one for the model in the large N limit.

Bethe Ansatz solutions of the $O(3)$ and $O(4)$ models have been obtained by solving equivalent fermionic models[16, 17]. Using these and relating them to the perturbative results a recent calculation by P. Hasenfratz *et al.* [18] calculated the exact value of the ratio of the mass gap of the model to the Λ parameter; for the O(3) model it is $m/\Lambda_{\overline{MS}} = 8/e$. This is an interesting

and important recent result. Because the applicability of the Bethe Ansatz has not been proven it is important to check this result with all methods possible. Monte Carlo simulation of the lattice version of the model offers the best prospects for such a test and has already played an important part in verifying the theoretical predictions of its properties.

Now, with new algorithms which update the important degrees of freedom of the theory directly and thus much more efficiently, it is possible to get around the problem of critical slowing down that previously plagued simulations near critical points. The overrelaxation algorithm and the more recent cluster algorithms allow the study of the model on large lattices, enabling precise measurements of correlation functions and physical quantities even for large correlation lengths, of the order of hundreds, unaffected by finite size errors. Thus a clear test can be made of the weak coupling expansion obtained using the Renormalization Group (in particular of asymptotic scaling of the mass gap and susceptibility of the model), and of the exact value of the mass gap.

The previous most precise Monte Carlo calculations to date with the Standard Action (of Wolff[13]) do not see asymptotic scaling even at correlation length of about 120; even the latest MCRG calculation[19] reported asymptotic scaling for the standard and tree-level improved actions (TIA) at a value of

$m/\Lambda_{\overline{MS}} = 3.4(1)$ which is different from the exact value. While there is interesting work using the $1/N$ expansion that compares with the exact result using numerical evaluation on finite lattices[20], since, for $N = 3$ the expansion parameter is not small, this expansion can obviously by itself provide conclusive evidence. Furthermore these results rely on numerical evaluation of large integrals (cost $\propto L^4$) and an extension from finite volume to infinite volume, and are thus very hard to compute for larger lattices; they are only available for $\beta \leq 2.1$, and thus although interesting cannot be used to extend our knowledge about asymptotic scaling much, even in an approximate way.

Thus Monte Carlo simulation and the Monte Carlo Renormalization group are the best methods that can be used to investigate non-perturbative aspects of the model and check asymptotic scaling. We performed a Monte Carlo simulation to obtain the infinite volume value of the correlation length and susceptibility. First, our simulations using the overrelaxation algorithm confirm the measurements of the susceptibility and mass gap of the model for $\beta \leq 1.9$, *i.e.*, $\xi \leq 120$, made in the cluster Monte Carlo study of Wolff. This is important, as no other study has checked the results for values of $\beta > 1.7$ because critical slowing down is severe for other algorithms; in particular, we verify the non-monotonic behavior of the ratio of the mass gap and susceptibility to the respective perturbative two-loop approximations.

Our results then go on to extend the region measured, first with the over-relaxation algorithm up to $\beta = 2.05$ on a lattice of length $L = 1024$ where $\xi \simeq 300$; subsequently we used a cluster Monte Carlo algorithm on a Connection Machine-2 (CM-2) to simulate at up to $\beta = 2.15$ on a lattice of length $L = 2048$, where $\xi > 500$.

We are able to extract the infinite volume correlation length for $\beta \leq 2.1$ and the susceptibility for $\beta \leq 2.0$. Our estimates are not consistent with asymptotic scaling for the mass gap and susceptibility of the model in this coupling constant range. In particular the value obtained at $\beta = 2.1$ for the ratio of the mass gap to $\Lambda_{\overline{MS}}$, the lambda parameter in the modified minimal subtraction scheme, is $m/\Lambda_{\overline{MS}} = 3.41(4)$, which is still 16% larger than exact result. This is the case even though the model is deep in the continuum limit, with $\xi = 420 \pm 5$. The susceptibility also shows no clear sign of attaining asymptotic scaling, growing faster than the three-loop equation predicts.

We also check on an assertion that using a rescaled coupling constant β_E, that is derived from the energy, improves the convergence to asymptotic scaling[21, 22]. This is equivalent to the coupling β in the limit $\beta \to \infty$ and effectively perform a resummation of the series that it is argued could take into account the disturbance from any nearby unphysical singularities[21]. We see that the two-loop results are much closer to the exact result of $m/\Lambda_{\overline{MS}}$.

However the third-loop correction, which we calculate, moves the results away from the exact results and towards that observed when using the usual coupling β. Thus the proximity of the two-loop results to the exact value must be attributed to coincidence.

Clearly there are other contributions to the β-function of the model that cause the difference between the predicted and observed values of the mass gap. These could be either higher order perturbative terms or non-perturbative effects. A possible source of non-perturbative effects is a pair of singularities in the complex coupling constant plane, observed by Butera *et al.* [23] and estimated to occur at $\beta = 1.9(1) \pm 0.3i$. Thus a search at higher values of β is of interest, since the influence of these singularities should be ameliorated.

Before we describe our subsequent calculations we will concern ourselves with the Monte Carlo update algorithms we use for these simulations. For the first part of this study, we used a hybrid simulation algorithm that combines overrelaxation and Metropolis updates. Most other local Monte Carlo update algorithms, suffer from critical slowing down: autocorrelation times τ diverge as a power $z > 2$ of the correlation length as the critical point of the model is approached ($\tau = c\xi^z$). Thus large simulations at large correlation lengths would require CPU time proportional to ξ^{d+z} (where d is the dimension of

the system), *i.e.*, ξ^4 in our case, which is prohibitively large.

The overrelaxation algorithm is a generalization of the successive overrelaxation algorithm for the iterative solution of matrix equations for the update of simple multi-quadratic actions. It is a local algorithm and for free fields[24] it obtains $z = 1$. For models that have additional interactions one of the ways to use the algorithm is to use it in its microcanonical variant and add some simple Metropolis to provide ergodicity. We already showed that, for the XY model, this hybrid overrelaxation method obtains significant speedup over other local algorithms[25] and, when the number of overrelaxation sweeps per metropolis sweep is held constant, it reduced the dynamical critical exponent in this model to between $z = 1.2$ and 1.48.

In our present study we observe that, by increasing the ratio of overrelaxation sweeps to metropolis sweeps in proportion to the correlation length ξ, the effort expended to obtain the same autocorrelation time rises with a smaller exponent, $z' = 1.08(1)$. Since this hybrid overrelaxation algorithm is the fastest currently known for other models, in particular for quenched lattice gauge theory, (and for other actions of this model that include antiferromagnetic interactions, like, e.g., the tree-level improved action) this result is of particular interest. Furthermore, our result has been confirmed in an analytical study of this algorithm for the Gaussian model by Wolff[26]. He

shows that a similar choice of the mix of overrelaxation sweeps and heat bath sweeps obtains optimal performance with $z = 1$. In a recent paper Bathas and Neuberger[27] showed that a class of generalizations of the overrelaxation algorithm are incapable of reducing z below its standard overrelaxed value of $z \sim 1$. They reach the tentative conclusion that there exists a barrier at $z \sim 1$ for local update algorithms.

It seems that the only way to obtain faster decorrelation is to use the cluster algorithms invented by Swendsen and Wang[28] for Potts models and extended to $O(N)$ models by Wolff[29]. These obtain very small decorrelation times for all correlation lengths, with exponents z that are so small they are hard to measure; in most cases z is between 0 and 0.3.

However in order to use cluster algorithms on large lattices we had to resolve some issues involved in adapting them to parallel computers. The reason is that parallel computers offer the most powerful computing resources for Monte Carlo simulation of large systems at the present time. Capable of a large number of floating point operations - of the order of 1-10 GFlops for many current machines, they typically have large memories of several gigabytes (Gbytes) and are cost effective relative to other high performance computers, e.g., vector machines like Crays. Current trends are such that they will continue to the best performance for the foreseeable future. They

are very well suited to local algorithms, like the single site Metropolis and overrelaxation, in which information of a (small) finite volume around a site is used to update the degree of freedom. The advent of cluster algorithms, produced a problem. In one step of the method the bonds of the Hamiltonian are replaced by a freezing bond or deleted, and after this all sites joined by frozen bonds are updated simultaneously. An essential part of the update procedure is thus to identify the clusters of sites joined by frozen bonds. This, however, is a non-local task; at criticality, the largest cluster percolates, so information must travel over all length scales up to the length of the lattice, and via fractal labyrinthine paths.

We address the particularly hard problem of obtaining a practical and simple algorithm for cluster identification within the data parallel model of parallel computation. This model is the most suitable for Single Instruction Multiple Data (SIMD) computers like the Connection Machine-2 and MasPar MP-1 and MP-2; algorithms for this model can also be adapted for Multiple Instruction Multi Data parallel computers like, e.g., the Ncube hypercubes and the Touchstone Delta. First we note that naive algorithms which use only local steps have very poor performance, since at criticality they require an average number of steps proportional to the length L of the lattice. We construct two algorithms to perform cluster labeling : the first is a simple

multi-scale algorithm that uses connection variables at lengths that are powers of two and a simple prescription that progressively sets more connections at each scale, thus accelerating the rate that information travels over large distances. The other uses the irregular communication routines to get and send cluster labels over ever increasing length scales.

For both methods the average number of iterations required to label configurations of bonds (at the critical point of the Ising model) scales with the logarithm of the length of the lattice. The time taken by the get/send algorithm is the best obtained for this problem on the CM-2.

Finally we perform a Monte Carlo renormalization group study of the model. This probes a region of the model not previously investigated and provides the best comparison with the exact result that can presently be obtained. To perform the calculation we have made the first extensive use of an innovative MCRG method proposed several years ago, one of few practical and general methods for extracting the couplings of the block Hamiltonians.

We find that the method performs well and can extract the block Hamiltonian for several blocking levels of the transformations we tried; its systematic errors are quantifiable - it has errors from the truncation of the trial Hamiltonian and from finite size effects. By studying the change of couplings under the standard (averaging) block transformation for the standard action

we see convergence to a renormalized trajectory after about four blocking steps. The same curve is obtained for an action proposed by Shenker and Tobochnik, who used simple block spin renormalization group arguments in an attempt to derive an action close to the renormalized trajectory. The trajectory is also close to that seen by Okawa using a very different method, but the two curves do not coincide; statistical errors however (not quoted for that work) and known systematic errors, which include truncation effects in both method and finite size effects (due to the older calculation's lattice size of 128^2) account for the differences.

We also studied the flow of couplings for a probabilistic block transformation. This was proposed by P. Hasenfratz *et al.* [30] as an improved alternative, in order to obtain a transformation whose renormalized trajectories is closer to the nearest neighbor Hamiltonian and thus hopefully is reached after fewer blocking steps. The block spin $\vec{b}$ is chosen from a probability distribution $P(\vec{b}) \propto e^{-C\vec{b}\cdot\vec{S}}$ where $\vec{S}$ is the sum of spins in a block and $C = C(\beta)$ is a tunable parameter. For the choice $C = c_\beta$ leading order perturbation theory suggests that $c = c_{opt} = 2.3$ is optimal for the O(3) model. We tried to find the renormalized trajectory for this transformation. Several points starting at low initial β appear to almost converge to a single curve; however, for larger initial couplings ($\beta \geq 2.0$), some of the couplings do not converge

even after five blocking steps. Thus this choice does not seem useful.

With another choice, $C = 5.0$, though, we are able to observe clear convergence to a single curve, which we identify as the renormalized trajectory of this transformation. Using this, we extract the discrete β function of the model $\Delta\beta(\beta)$ by comparing the results of simulations on different lattice sizes. (The discrete beta function is defined by $\Delta\beta(\beta) = \beta' - \beta$ where β' is chosen so that $\xi(\beta') = \frac{1}{2}\xi(\beta)$). We obtain agreement with the measurements of Hasenfratz and Niedermayer at $\beta = 2.14$ and 2.26 and are able to measure up to $\beta = 2.62$ where the correlation length $\xi = 9180 \pm 100$. A comparison to the perturbative three-loop results for $\Delta\beta$ shows that for $\beta \leq 2.5$ our results slowly converge to the prediction, with the last two being compatible with the three-loop equation within statistical errors; however our estimate at $\beta = 2.62$ is much larger than the three-loop expectation. Thus our results do not demonstrate asymptotic scaling within the region studied.

Furthermore, if we use our measurements of the discrete β function along the chain of points $(\beta = 2.02, 2.14, \ldots, 2.50, 2.62)$ and that of the correlation length at $\beta = 2.0$ and 2.05, we can estimate the correlation length at each of these subsequent points. (This was the method used to obtain the estimate of $\xi(\beta = 2.62) = 9200$.) From these we can obtain the value $m/\Lambda_{\overline{MS}}(\beta = 2.62) = 3.31(4)$. This remains compatible with the value

$m/\Lambda_{\overline{MS}} = 3.4(1)$, measured for the TIA by Hasenfratz and Niedermayer, where asymptotic scaling for the $\Delta\beta$ function is seen within statistical errors. Thus universality appears to hold. However for the standard action - as the equivalent measurement of the discrete beta function already showed - the value of $m/\Lambda_{\overline{MS}}$is not constant. Its value remains 10% higher than the exact value obtained from the Bethe Ansatz solution, showing that corrections to the three-loop formula for m (either higher order terms or non-perturbative effects) must be higher than expected.

Using however the alternative coupling constant β_E we see that for the four points $\beta = 2.14$ to 2.50 the values of $m/\Lambda_{\overline{MS}}$ we get by dividing by the two- (or the three-) loop equations that agree with each other. This lead us to provisionally conclude that asymptotic scaling in this coupling may have been observed. A problem with this is that the last point, at $\beta = 2.62$, disagrees with these values; however, this point might be off because of a combination of statistical and systematic effects. If this objection is put aside for a moment we can examine the value we get for $m/\Lambda_{\overline{MS}}$. To three-loops this is $m/\Lambda_{\overline{MS}} = 3.047(35)$ at $\beta = 2.50$, which is very close to the exact value $m/\Lambda_{\overline{MS}} = 2.943$; the difference is small enough that the next order correction could account for it. Thus it would appear that credence is lent to the assertion[21] that the alternative choice of coupling β_E will achieve

asymptotic scaling much faster than the original coupling β.

We will now describe the organization of the subsequent chapters. First, we will review some of the important treatments of the model that we have mentioned, that lead to the predictions of its properties, in chapter 2. We will then present the new monte Carlo simulation techniques that allow us to do high accuracy simulations of the model at large correlation lengths in chapter 3. Next we will describe our Monte Carlo simulations on large lattices to test asymptotic scaling of the mass gap and susceptibility in Chapter 4. A description of our algorithms for labeling clusters of spins and detailed results for these are presented in Chapter 5. Our Monte Carlo renormalization group calculation is in Chapter 6.

Chapter 2

Theory of the Model

The O(N) nonlinear sigma model and its lattice regularized counterpart, the classical Heisenberg model, have been extensively studied.

Their behavior in two dimensions is dictated by the lack of spontaneous magnetization for models with a continuous symmetry for this number of dimensions, in accordance with the Mermin-Wagner-Coleman theorem[31, 32]. Their long wavelength behavior has been determined using perturbation theory in the low temperature limit, and the behavior of the mass gap (correlation length) and susceptibility of the theory have been determined using the Renormalization Group.

The models are believed to be asymptotically free for $N > 2$ and possess a dynamically generated mass gap; for $N = 3$ the model also has instanton

solutions, and both these properties are similar to those of non-Abelian gauge theories in four dimensions.

The case of $N = 2$ is the XY model, which has very different behavior driven by topological effects[33]: it has an infinite order phase transition driven by a condensate of vortices which produces an infinite correlation length on the low temperature side of the transition, and a diverging correlation length in the high temperature phase, which is a phase of dilute vortices.

In contrast, the O(N) model for $N > 2$ has only one phase, the massive phase of exponential correlation functions that is seen in the high temperature/strong coupling expansions of the model[2]. These predictions are made by the Low Temperature Renormalization Group of Polyakov[1]; we will review them in this chapter.

Similarly to the Massive Thirring Model, the O(N) model has an infinite set of conservation laws which lead to conservation of particle number and factorizability of the S matrix. Using this an exact S matrix has been proposed for the model. An exact solution of the model has also been obtained.

We also examine the consequences of knowing the exact solution and the S matrix, in particular in combining it with the perturbative solution, to get the exact value for the ratio of the mass gap to the Λ parameter.

2.1 The connection between lattice and continuum models

We start by examining the relation between the continuum model and its lattice regularized counterpart. This is the two-dimensional classical Heisenberg model, or O(N) invariant Heisenberg model; the Hamiltonian of the model is

$$H = -\sum_{i,j} V_{ij}\ \vec{s}_i \cdot \vec{s}_j \qquad (2.1)$$

where $\vec{s}_i$ is an N component vector of unit length and $V_{ij} = V(i-j)$ is a translation and rotation invariant potential. The partition function of the model is

$$Z = \int \prod_i d\mu(\vec{s}_i)\ e^{-\frac{1}{T}H} \qquad (2.2)$$

where T is a dimensionless temperature and $d\mu()$ denotes the $O(N)$ invariant measure.

The constraint $(\vec{s}_i)^2 = 1$ means that the components of the spins $\vec{s}_i$ are not independent degrees of freedom. In order to study the model we identify the $N-1$ (independent) degrees of freedom by expanding the spin about a direction $\hat{u}$:

$$\pi_i^a = s_i^a - (\vec{s}_i \cdot \hat{u})\ u^a \qquad (2.3)$$

for $a = 1, \ldots, N-1$ and

$$\sigma = \vec{s}_i \cdot \hat{u} = \sqrt{1 - \vec{\pi}^2}. \tag{2.4}$$

For dimensions other than two, $\hat{u}$ can be thought of as the direction of spontaneous magnetization. The partition function thus becomes

$$Z = \int \prod_i \frac{d\vec{\pi}_i}{[1 - \vec{\pi}_i^2]^{1/2}} \exp\Big(\frac{1}{T} \sum_{ij} V_{ij}[\vec{\pi}_i \vec{\pi}_j + (1 - \vec{\pi}_i^2)^{1/2}(1 - \vec{\pi}_j^2)^{1/2}]\Big). \tag{2.5}$$

After expanding the Hamiltonian in powers of π^2, a series in terms of loops can be obtained. This gives an expansion in the temperature T, in which the propagator involved is the inverse of the two-point Green's function

$$G_{ab}(q) = \frac{T}{\tilde{V}(0) - \tilde{V}(q)} \delta_{ab}, \tag{2.6}$$

where $\tilde{V}(q) = \sum_j V_{ij} e^{\vec{q} \cdot \vec{r}_{ij}}$ is the Fourier transform of the potential.

The interactions in the model originate in the higher order terms of the expansion of the square roots in the σ terms and from the integration measure

$$\prod_i \frac{1}{\sqrt{1 - \vec{\pi}_i^{\,2}}} = \exp\Big[-\frac{1}{2} \sum_i \ln(1 - \vec{\pi}_i^{\,2})\Big]. \tag{2.7}$$

Since we are interested in the long distance limit of the model we eliminate all but the most divergent parts of the propagator to $G_{ab} \simeq \frac{T}{q^2}\delta_{ab}$. This simplification makes the diagrammatic expansion the same as that of the nonlinear sigma model, with the action

$$\mathcal{A} = \int dx \left[(\nabla \vec{\pi}(x))^2 + \left(\nabla \sqrt{1 - \vec{\pi}(x)^2} \right)^2 \right]. \tag{2.8}$$

In order to study this model the propagator's infrared divergences in two dimensions must be regulated. There are two approaches to this; the more recent one[34, 35] relies on Elitzur's theorem on the vanishing of infrared divergences for Green's functions that are invariant under the symmetry of the group[36]. Using it one studies only products of the fields like the two-point function and four-point functions.

The original approach of Brezin and Zinn-Justin[4] uses a magnetic field H along the Nth axis as a regulator. Thus the O(N) symmetry of the model is broken, and is recovered only in the limit of a vanishing field $H \to 0$. The additional term $-H\sigma(x)$, which is introduced, can in turn be expanded in powers of $\pi(x)$ and the two-point Green's function becomes

$$G_{ab}(q) = \frac{T}{q^2 + H} \quad \delta_{ab}. \tag{2.9}$$

The expansion of Z can then be carried out in $d = 2 + \epsilon$ dimensions. The terms coming from the measure vanish in dimensional regularization. They have the effect of preserving the O(N) invariance of the theory by canceling the mass terms that would otherwise be induced by the breaking of this symmetry, which also vanish in this regularization.

The four-point interaction can be read off from the action, and in momentum space is

$$\frac{1}{8T} \sum_{q,q_1,q_2} (q^2 + H)\pi^{\alpha}_{q_1} \pi^{\alpha}_{q-q_1} \pi^{\beta}_{q_2} \pi^{\beta}_{q-q_2}. \tag{2.10}$$

Six and more point interactions are also generated, and can similarly read off. However the behavior for two dimensions remains problematic, e.g., the two-point vertex function $\Gamma^{(2)}_{ab}(q) = \delta_{ab}\Gamma^{(2)}(q)$ calculated to one loop is infinite

$$\Gamma^{(2)}(q) = \left(p^2 + \frac{H}{T}\right) - \frac{1}{\epsilon}\, H^{\epsilon/2} \left[p^2 + \frac{N-1}{2} H\right] + O(\epsilon, T). \tag{2.11}$$

2.2 The Renormalization Group treatment

This pole of the two-point vertex function in exactly two dimensions is a consequence of ultraviolet divergences, and must be rectified by renormalization. The symmetry of the model insures that only two renormalization constants

are required. An arbitrary mass scale μ is introduced and the renormalized quantities are the dimensionless rescaled temperature t

$$T = \mu^{-\epsilon} Z_t t \tag{2.12}$$

and the field π_R^a:

$$\pi^a = Z_\pi^{1/2} \pi_R^a \tag{2.13}$$

The magnetic field can absorb the field renormalization and its renormalized value is

$$H = Z_t Z_\pi^{-1/2} h \tag{2.14}$$

The renormalized Greens and vertex function will thus become

$$G_{\pi R}^{(M)}(p_i; t, h, \mu) = Z_\pi^{M/2} G_\pi^{(M)}(p_i; \mu^\epsilon Z_t^{-1} T, Z_t^{-1} Z_\pi^{-1} H, \Lambda) \tag{2.15}$$

$$\Gamma_{\pi R}^{(M)}(p_i; t, h, \mu) = Z_\pi^{-M/2} \Gamma_\pi^{(M)}(p_i; \mu^\epsilon Z_t^{-1} T, Z_t^{-1} Z_\pi^{-1} H, \Lambda) \tag{2.16}$$

where Λ is a reference scale.

The right-hand side of these equations is expanded order by order in t and must be finite at each order in the limit of $\epsilon = 0$. Expanding Z_t and Z_π

and substituting into equation (2.11) leads to

$$Z_t = 1 + \frac{N-2}{\epsilon} t + O(t^2) \tag{2.17}$$

$$Z_\pi = 1 + \frac{N-1}{\epsilon} t + O(t^2) \tag{2.18}$$

This allows us to extract the β function, that determines the flow of the coupling constant with change of scale, and the anomalous dimension γ of the π field from the variation in the mass μ,

$$\beta(t,\epsilon) = \quad \mu \tfrac{\partial t}{\partial \mu}|_T \quad = \epsilon t - (N-2)t^2 - (N-2)t^3 + O(t^4) \tag{2.19}$$

$$\gamma_\pi(t) = \mu \tfrac{\partial \ln Z_\pi}{\partial \mu}|_{\text{Bare}} = -(N-1)t + O(t^3) \tag{2.20}$$

In the above equations we have also used the two-loop results[4].

The connected correlation functions fulfill the Renormalization Group equation

$$\left\{ \mu \frac{\partial}{\partial \mu} + \beta(t) \frac{\partial}{\partial t} + \frac{1}{2} N \gamma_p i(t) + \left(\frac{1}{2} \gamma_p i(t) + \frac{\beta(t)}{t} - \epsilon \right) h \frac{\partial}{\partial h} \right\} G_{\pi R}^{(M)}(p_i; t, h, \mu) = 0 \tag{2.21}$$

Although the expansion used is not good in two dimensions, the renormalization group functions $\beta(t,\epsilon)$ and $\gamma_\pi(t)$ are regular at $\epsilon = 0$ and are thus

valid. Thus, from the beta function of equation (2.19) we see that in two dimensions the theory is asymptotically free for $N > 2$, like non-Abelian gauge theories in four dimensions. and, since the coefficients of the renormalization group expansion are regular at $\epsilon = 0$, the behavior in two dimensions can be extracted. The O(N) symmetry is restored and the field becomes massive. The correlation length ξ is determined by

$$\left(\mu\frac{\partial}{\partial\mu} + \beta(t)\frac{\partial}{\partial t}\right)\ \xi(t,\mu) = 0. \tag{2.22}$$

Given the dimensional argument $\xi \sim \mu^{-1}$, can be integrated easily to determined the mass gap of the theory $m = 1/\xi$

$$m(t) = \mu\ t^{-\frac{1}{N-2}} \exp\left(-\frac{1}{t(N-2)}\right) f(t) \tag{2.23}$$

where $f(t)$ is a function regular at $t = 0$. Perturbation theory cannot determine the value of $f(0)$.

The susceptibility of the theory is also determined:

$$\chi(t) \sim t^{-1-\frac{3}{N-2}} \exp\left(-\frac{2}{t(N-2)}\right) \tag{2.24}$$

2.3 The S matrix

The nonlinear O(N) sigma model in two dimensions has an infinite set of non-local quantum conservation laws[37]. These lead to conservation of particle number, *i.e.*, forbid particle creation, and thus lead to the factorizability of the S matrix into combinations of two particle S matrices.

A.B.Zamolodchikov *et al.* [15] determined the two particle S matrices that are compatible with the hypotheses that the elementary particles of the model are massive, do not form bound states and belong to vector O(N) multiplets. Conservation of two-momentum then lead to a simple two particle S matrix

$$
\begin{aligned}
&{}_{ik}S_{jl}(p_1, p_2, p_1', p_2') = \\
&\delta(p_1 - p_1')\delta(p_2 - p_2')\left(\delta_{ik}\delta_{jl}\sigma_1(s) + \delta_{ij}\delta_{kl}\sigma_2(s) + \delta_{il}\delta_{jl}\sigma_3(s)\right) \\
&+(p_1' \leftrightarrow p_2')
\end{aligned}
$$

where $s = (p_1 + p_2)^2$. Using the rapidities θ_i of the particles to replace the momenta, $p_i^0 = m\cosh\theta_i$ and $p_i^1 = m\sinh\theta_i$, where m is the mass, simplifies the treatment. The Mandelstram variable s can then be replaced by $\theta = \mid \theta_1 - \theta_2 \mid$ as $s = 2m^2(1 + \cosh\theta)$. The functions $\sigma(\theta) = \sigma(s(\theta))$ are

related by the unitarity and analyticity of the S matrix,

$$\sigma_2(\theta)\sigma_3(-\theta) + \sigma_2(-\theta)\sigma_3(\theta) = 0 \tag{2.25}$$

$$\sigma_2(\theta) = \sigma_2(i\pi - \theta) \tag{2.26}$$

$$\sigma_3(\theta) = \sigma_1(i\pi - \theta) \tag{2.27}$$

and two other similar equations. Furthermore the factorization of the S matrix into sums of two-particle terms requires certain relations between the σ functions. These self-consistency equations ensure the coherence of outgoing waves, that is required in order that their monochromatic nature is maintained. The first of these three equations

$$\sigma_2(\theta)\sigma_3(\theta+\theta')\sigma_3(\theta') + \sigma_3(\theta)\sigma_3(\theta+\theta')\sigma_2(\theta') = \sigma_3(\theta)\sigma_2(\theta+\theta')\sigma_3(\theta')$$

allows the elimination of $\sigma_3(\theta)$ up to a parameter λ

$$\sigma_3(\theta) = -i\frac{\lambda}{\theta}\sigma_2(\theta)$$

which together with (2.27) determine σ_1 in terms of σ_2. With the substitution of σ_1 and σ_3 two other equation determine the parameter λ : $\lambda = \frac{2\pi}{N-2}$.

Finally equation (2.25) leads to an equation for σ_2

$$\sigma_2(\theta)\sigma_2(-\theta) = \frac{\theta^2}{\theta^2 + \lambda^2}$$

This and equation (2.26) must be satisfied by $\sigma_2(\theta)$, leading to a general solution

$$\sigma_2(\theta) = \left[\prod_{k=1}^{M} \frac{\sinh\theta + i\sin a_k}{\sinh\theta - i\sin a_k}\right] \sigma_2^0(\theta)$$

where a_k are M arbitrary real constants and $\sigma_2^0(\theta)$ is the solution with the least number of singularities. For $N = 3$ this is

$$\sigma_2^0(\theta) = \frac{\theta\ (i\pi - \theta)}{(2i\pi - \theta)(i\pi + \theta)}$$

This solution is the only one whose spectrum has no isospin degeneracy. It also has no poles on the physical part of the s-plane, so there are no physical bound states.

Thus using the unitarity, analyticity and factorization of the S matrix we have constructed a factorized S-matrix with $O(N)$ isotopic symmetry, for $N > 3$, by the bootstrap program. This has been compared with the $1/N$ expansion of the O(N) model showing that they are in agreement up to order $1/N$ originally[38] and subsequently[39] up to order $1/N^2$.

2.4 Bethe Ansatz solution of the O(3) model

Wiegmann[16] provided the exact solution of the O(3) nonlinear sigma model using the Bethe Ansatz technique. The model is considered with a source $\vec{h}_{mu}$ coupled to the Noether current $\vec{J}_\mu = \vec{s} \times \partial_\mu \vec{s}$.

The goal is to determine the free energy of the model $f(h)$

$$e^{-f(h)} = \int \mathcal{D}s(x) \exp\left(-A(s) - \vec{h}_\mu \cdot \int d^2x \vec{J}_\mu\right), \tag{2.28}$$

where $h^2 = \vec{h}_\mu \cdot \vec{h}_\mu$.

Polyakov and Wiegmann[17] solved the O(4) nonlinear sigma model by using the relation to the $SU(2) \otimes SU(2)$ model and solving an equivalent fermionic model. A generalization of this method can be applied to solve the equivalent fermionic model of the O(3) model. The exact solution is provided by a hierarchy of Bethe Ansatze that yields the ground state energy. From this the free energy difference of states in an external field is determined to be

$$f(h) - f(0) = -\frac{1}{2\pi} m \int_{-B}^{B} \cosh\theta \cdot \epsilon(\theta) d\theta \tag{2.29}$$

where $\epsilon(\theta)$ derived from the integral equation

$$\epsilon(\theta) - \int_{-B}^{B} \frac{1}{\pi^2 + (\theta - \theta')^2} \epsilon(\theta') d\theta' = h - m \cosh \theta \tag{2.30}$$

with B determined by the condition $\epsilon(\pm B) = 0$.

2.5 The exact value of the mass gap

Hasenfratz, Maggiore and Niedermayer [18] obtained the exact value of $m/\Lambda_{\overline{MS}}$ by comparing the Bethe Ansatz solution of the O(3) model of Wiegmann[16] with the result of weak coupling perturbation theory.

In particular, they consider the model in the presence of an external field h coupled to the Noether charge Q_3. Using perturbation theory they obtain an expansion of the free energy difference $f(h) - f(0)$ in terms of $\Lambda_{\overline{MS}}$ for the limit of $h \gg m$:

$$f(h) - f(0) = -\frac{h^2}{4\pi}\Big(\ln \frac{h}{e^{1/2}\Lambda_{\overline{MS}}} + \ln \ln \frac{h}{\Lambda_{\overline{MS}}} + O(1)\Big) \tag{2.31}$$

We then consider equation (2.29) from the Bethe Ansatz solution. We concentrate on the threshold $h \gtrsim m$ and the limit $h \gg m$. In the former a

simple iterative solution is possible giving:

$$f(h) - f(0) = -\frac{2}{3}\frac{\sqrt{2m}}{\pi}(h-m)^{3/2} \tag{2.32}$$

To check this one can observe that the proposed S-matrix is, in this limit, equivalent to that of a gas of impenetrable bosons in one dimension. From this one can obtain expression for the same free energy difference; the result agrees with the previous equation.

In the limit of large h, an expansion in terms of h/m with a multiplying factor of h^2 is expected, similar to equation (2.31). Assuming this form , a simple iterative calculation gives an approximate value for the constant term, and thus $\Lambda_{\overline{MS}}$. In contrast, considerable effort is required to prove this expansion by solving the integral equation in this limit using generalized Wiener-Hopf technique. The expansion obtained agrees with that of perturbation theory and the comparison reveals the mass gap m in terms of the Λ parameter (to within 10^{-8})

$$\frac{m}{\Lambda_{\overline{MS}}} = \frac{8}{e} \tag{2.33}$$

This is an important theoretical prediction. If either the Bethe Ansatz

solution or the proposed S matrix is correct, this calculation means that the renormalization group determines the behavior of the mass gap of the theory, in the low temperature limit, with no free parameters. Furthermore, since this theory is much easier to simulate than gauge theory in four Euclidean dimensions, a more stringent test of the results of the Renormalization group is possible.

It is with the methods we will use to conduct this test that we will concern ourselves next.

Chapter 3

Monte Carlo Algorithms

In the absence of analytical tools that evaluate most physical quantities of interest in the non-perturbative region, a powerful numerical method has been used on many problems: the Monte Carlo method. This tool is used in conjunction with analytical techniques applicable to special cases and limits. (There are many introductions to Monte Carlo, e.g., Binder and Heermann [40], and Sokal[41]).

In order to discuss with more clarity, we will use a typical problem from Statistical Mechanics. We introduce a field $S(x)$ on a d dimensional space with a Hamiltonian $\mathcal{H}$. For a typical quantity one wishes to measure $\mathcal{O}\{S(x)\}$, the average of interest can be expressed as

$$< \mathcal{O} >= \frac{\int \mathcal{D}S \ \mathcal{O} \ \exp -\beta\mathcal{H}(S)}{\int \mathcal{D}S \ \exp -\beta\mathcal{H}(S)}, \tag{3.1}$$

where $\mathcal{D}S = \prod_x d\mu(S(x))$ and $d\mu(S(x))$ is the invariant measure of the field $S(x)$.

In most cases people use *importance sampling* to generate a large number of different configurations S, using many different methods, that are distributed according to the probability distribution

$$P_{eq}(S) \equiv \pi_S = \frac{1}{Z} \exp[-\beta\mathcal{H}(S)] \tag{3.2}$$

where the normalization factor Z is the partition function:

$$Z = \int dS \exp[-\beta\mathcal{H}(S)] \tag{3.3}$$

If N configurations S_i are generated and N_0 are left out to allow the system to thermalize, the estimate of the average is

$$< \mathcal{O} >= \frac{1}{N - N_0} \sum_{i=N_0+1}^{N} \mathcal{O}(S_i) \tag{3.4}$$

3.1 Dynamical Monte Carlo

The dynamical Monte Carlo method is a widely used and general method for evaluating large dimensional integrals by importance sampling. It is a stochastic process and encompasses a large variety of techniques, which share a number of properties: starting from a configuration $\{S_0(x)\}$ of the system a Markov chain of configurations $\{S_i(x)\}$ is generated using a *transition rule* $\mathcal{T}$ to generate each one from the previous one:

$$\{S_i(x)\} \xrightarrow{\mathcal{T}} \{S_i'(x)\}$$

$\mathcal{T}$ is almost always probabilistic, and thus can be specified using the probability distribution $W(\phi \rightarrow \phi')$. In order that the configurations generated have a limit probability distribution, and that is the equilibrium one π_ϕ, it is enough that the transition procedure $\mathcal{T}$ satisfies two simple conditions,

- *Ergodicity.* Starting from any configuration S there exists a possible path to any other configuration S'. More precisely there exists a finite number of steps $n > 0$ for which the probability $P_n(S \rightarrow S')$ is finite. This is more precisely called *irreducibility.*

- *Stationarity.* For all configurations r,

$$\sum_x \pi_q \mathcal{T}_{qr} = \pi_r. \tag{3.5}$$

If these conditions hold then we are guaranteed that the probability distribution of configurations generated will approach the desired stationary distribution in the limit of infinite number of steps: The infinite time averages of quantities measured will tend to the ensemble averages.

The stationarity condition is usually achieved by satisfying another more restrictive condition: *detailed Balance* :

$$\pi_S P(S \to S') = \pi_{S'} P(S' \to S)$$

Autocorrelation: Since every configuration S' is generated starting from a previous one S, it will contain memory of it. A "essentially new" configuration, statistically uncorrelated with S will only be produced after a certain number of Monte Carlo steps. The measure of the number of steps to a "new" configuration is the decorrelation time τ.

3.2 The Metropolis procedure

The procedure proposed by Metropolis[42] in 1953 and extended slightly by Hastings[43] is a very general procedure that governs almost all Monte Carlo simulations. A change is proposed according to an arbitrary rule and is accepted according to a simple criterion. Let $P^{(0)}$ is arbitrary irreducible transition matrix, called the *proposal* matrix. Each transition $\phi \rightarrow \chi$ is then accepted with probability $a_{\phi\chi}$ or rejected otherwise, leaving a trivial transition $\phi \rightarrow \phi$.

In order to satisfy detailed balance it is necessary and sufficient that the probabilities $a_{\phi\chi}$ follow:

$$\frac{a_{\phi\chi}}{a_{\chi\phi}} = \frac{\pi_\chi p^{(0)}_{\chi\phi}}{\pi_\phi p^{(0)}_{\phi\chi}} \tag{3.6}$$

for all $\phi \neq \chi$. This can easily be done by choosing

$$a_{\phi\chi} = F\left(\frac{\pi_\chi p^{(0)}_{\chi\phi}}{\pi_\phi p^{(0)}_{\phi\chi}}\right) \tag{3.7}$$

where $F : [0, +\infty] \rightarrow [0, 1]$ is a function for which for all x

$$\frac{F(x)}{F(1/x)} = x \tag{3.8}$$

3.2.1 The single-site 'Metropolis algorithm'

The most common choices of the very general Metropolis procedure are referred as the "Metropolis algorithm." In many cases the proposal matrix $P^{(0)}$ is symmetric, *i.e.*, the $\pi_\chi p^{(0)}_{\chi\phi} = \pi_\phi p^{(0)}_{\phi\chi}$. It thus cancels out in equations (3.6) and (3.7). α is usually adjusted to give an average acceptance of 50%.

The most popular choice is making the proposal a change in the field at one site from ϕ to a nearby value $\phi + \alpha\eta$, where α is an adjustable parameter and η a unit Gaussian random variable.

3.2.2 The Heat-Bath algorithm

For many actions a very different approach is possible: Considering a single site, keeping all other sites fixed and consider the conditional probability distribution generated by the action as a distribution from which to pick a new element. Thus the previous value of the spin at that site is ignored. For example, in the O(3) model the new spin is chosen from the distribution

$$P^\pi(s_i|\{s_j\}, j \neq i) = A \cdot \exp\left(\beta \vec{s_i} \cdot \sum_{<j,i>} \vec{s_j}\right) \tag{3.9}$$

where A is a normalization constant .

3.2.3 Critical slowing down

The single-site standard Metropolis and the heat bath algorithms provide a simple way to simulate many statistical systems. However they suffer from an important problem when used to study systems near criticality. As the correlation length of the system grows, the decorrelation time increases proportionately:

$$\tau = c \cdot \xi^z \tag{3.10}$$

where z is the dynamical critical exponent of the algorithm. The reason is that near the critical point spins are correlated over distances of the order of the correlation length ξ. Since these algorithms are local, information on a change will perform a random walk, and in 'time' t travels distance proportional to $\sqrt{t}$. Thus in order for a change to be made on the scale of ξ, it will take 'time' roughly ξ^2. Typically $z \geq 2$.

If this cannot be circumvented it will severely limit the ability to simulate close to critical points, *i.e.*, at large correlation lengths. In particular, as, with any algorithm, the time taken to do one sweep (update of all the sites of a lattice) is proportional to the volume, and the number of sweeps to generate a statistically independent configuration rises as ξ^z with $z \geq 2$, the computational cost rises very rapidly - like $\xi^z L^d \simeq \xi^{d+z}$. with the system

size L and correlation length ξ.

Thus algorithmic improvements that reduce the value of z are necessary in order to simulate large systems near criticality.

3.3 The overrelaxation algorithm

The first of these, the overrelaxation algorithm was invented by Adler[44] by generalizing the successive overrelaxation methods for solving systems of linear equations to stochastic processes for multi-quadratic actions. This algorithm was similar to the heat-bath algorithm; a new field was chosen from a probability distribution that is a function of the spins that field is coupled to. After identifying the minimum of the action we find the point opposite to the original value from the previous value of the field (and equidistant from the minimum), which we call the mirror point. The central point of the distribution is chosen at a point between the minimum and the mirror point. A new adjustable parameter ω, analogous to that of successive overrelaxation, determines both the location of the central point and the width of the distribution .

For the multi-quadratic action

$$S(x_i, \{x_j, j \neq i\}) = a(x_i - x_i^0)^2 + c \tag{3.11}$$

with a, x_i^0 and c functions of $\{x_{j\neq i}\}$, detailed balance is satisfied by the choice

$$x_i' = x_i^0 + (1-w)x_i + w(2-w)a^{-1/2}\eta \tag{3.12}$$

where η is unit Gaussian noise.

The values ω can take are thus restricted: $0 < \omega < 2$. Outside this range the algorithm diverges. In the limit that $\omega \to 2$ the algorithm becomes deterministic and microcanonical, always choosing the mirror point while for $\omega = 1$ the heat-bath algorithm, while in values in between we obtain successive overrelaxation.

Adler's overrelaxation algorithm has been shown to have significantly reduced critical slowing down for the Gaussian model[24], to $z = 1$ for both the case of sequential updating and of checkerboard updating. For this performance, it is necessary to tune the parameter w appropriately.

For other models, which do not have a multi-quadratic action, the simple equation determining the probability distribution is no longer applicable. The correct probability distribution is more complicated. Therefore there is a choice of approaches. Either the probability distribution must be used, exactly or approximately using an accept/reject step, or the microcanonical limit can be used which must be combined with another update procedure

- needed to provide ergodicity. The latter, hybrid, approach is the one we have chosen. This provides a very simple and fast prescription for the key part of the update procedure which is easily vectorizable and parallelizable.

In particular in our implementation we partitioned the lattice in one-dimensional slices among the processors. We used an updating scheme in which several rows of reds spins were updated simultaneously, followed by the lower neighboring rows until all red spins had been updated, and then all black spins were updated in a similar manner.

3.4 Cluster algorithms

A very different method has been discovered recently that works to alleviate critical slowing down even further. It is non-local, and is based on a correspondence between the Potts Model and the random percolation model long ago discovered by Fortuin and Kasteleyn[45]. Swendsen and Wang[28] invented this algorithm for a ferromagnetic Potts model, and it has since been extended, principally for $O(N)$ models by Wolff[29]

In order to talk about the extension to the O(N) models, we will briefly describe the Potts model version[28]. We define this for a spin σ which take

the value 1 to q with the Hamiltonian:

$$\mathcal{H} = K \sum_{<i,j>} (\delta_{\sigma_i,\sigma_j} - 1) \tag{3.13}$$

Starting from the partition function

$$Z = \mathrm{Tr}_{\{\sigma\}} e^{\mathcal{H}} \tag{3.14}$$

they consider the interaction between two site l and m and removè it from the Hamiltonian:

$$\mathcal{H}_{(l,m)} = K \sum_{<i,j> \neq <l,m>} (\delta_{\sigma_i,\sigma_j} - 1) \tag{3.15}$$

and the partition function can be split into two parts

$$Z = Z^{same}_{<l,m>} + e^{-K} Z^{diff}_{<l,m>} \tag{3.16}$$

where the spins at l and m have been restricted to have the same

$$Z^{same}_{<l,m>} = \mathrm{Tr}_{\{\sigma\}} \exp(\mathcal{H}_{(l,m)}) \delta_{\sigma_i,\sigma_j} \tag{3.17}$$

and different values

$$Z^{diff}_{<l,m>} = \mathrm{Tr}_{\{\sigma\}} \exp(\mathcal{H}_{(l,m)})(1 - \delta_{\sigma_i,\sigma_j}) \tag{3.18}$$

We can now define a partition function in which the term $< l, m >$ does not exist, and the spins are independent:

$$Z^{indep}_{<l,m>} = \mathrm{Tr}_{\{\sigma\}} \exp(\mathcal{H}_{(l,m)}) = Z^{same}_{<l,m>} + Z^{diff}_{<l,m>} \tag{3.19}$$

and in terms of this our original partition function is

$$Z = (1 - e^{-K}) Z^{same}_{<l,m>} + e^{-K} Z^{indep}_{<l,m>} \tag{3.20}$$

If this procedure is used to eliminate all interactions one by one, one ends up with the partition function of the independent bond percolation model, with bond probability $p = (1 - e^{-K})$.

Swendsen and Wang used this to define a procedure for updating a Potts model by using the following prescription:

- All terms in the Hamiltonian are considered and bonds are created between sites that have the same spin value with probability p.
- The sites which are connected by bonds are identified as a cluster

- The previous values of the Ising variables are thrown out, and each cluster assumes a random value of the Potts spin

Thus a new configuration of the model has been generated. This procedure obeys detailed balance, as can be shown by considering the transition from one Ising configuration to another through each possible bond configuration separately.

The extension of this algorithm to the O(N) model was made by Wolff[29], effectively by embedding Ising spins in the N component spin by using

$$\vec{\sigma}_x = \vec{\sigma}_x^{\perp} + \varepsilon_x \mid \vec{\sigma}_x \cdot \vec{r} \mid \vec{r} \tag{3.21}$$

where $\vec{r}$ is a unit N-vector, $\varepsilon_x = \pm 1$ and $\vec{\sigma}_x^{\perp} = \vec{\sigma}_x - \mid \vec{\sigma}_x \cdot \vec{r} \mid \vec{r}$ is the component of $\vec{\sigma}_x$ perpendicular to $\vec{r}$.

This method was studied numerically for the XY and O(4) models in two dimensions[46], and it was shown that critical slowing down is (almost) completely eliminated, with z about 0.1 for the XY model.

A further contribution was the proposal to modify the cluster algorithm by building only one cluster at a time and always flipping it. Thus all the moves, rather than only half, will change the lattice, and decorrelation should be faster. This algorithm lies in the same universality class as the Swendsen-

Wang dynamics for many models , but the decorrelation time is halved.

Chapter 4

Large Volume Monte Carlo

4.1 Introduction

Monte Carlo studies of the two-dimensional O(3) spin model, which is the discretized $1+1$-dimensional O(3) nonlinear sigma model, have been pursued extensively for about ten years. An important reason for this is that in $1+1$ dimensions O(N) models with $N \geq 3$ are asymptotically free[1, 4], just as gauge theories like QCD are in $3+1$ dimensions. Thus by investigating the simpler O(3) model one hopes to understand critical behavior similar to that of QCD. Moreover, the O(3) model also possesses instanton solutions – another feature in common with QCD.

A consequence of asymptotic freedom is that, at weak enough coupling,

all physical quantities will scale according to the two-loop β function; this is called asymptotic scaling. It is the main goal of the Monte Carlo simulations of the O(3) model to seek to test and hopefully demonstrate asymptotic scaling. This however has proven to be an elusive goal. Early studies of the standard, nearest neighbor, action (SA) using standard Monte Carlo techniques on small lattices[3, 8, 9, 10] (*i.e.*, $L \leq 100$), and Monte Carlo Renormalization Group (MCRG) methods[11, 12], failed to demonstrate asymptotic scaling. Berg *et al.* [12] simulated a tree-level improved action (TIA), which includes next-nearest neighbor interactions, and a 1-loop improved action, with quadratic and quartic interactions up to a distance two, and observed behavior closer to the asymptotic behavior, but due to large statistical errors and the use of small lattices could not provide conclusive evidence.

Analysis of high temperature expansions have also addressed the problem. P. Butera *et al.* [23] explain the lack of asymptotic scaling for the susceptibility in terms of a pair of complex singularities in $\chi(\beta)$ near the real axis. Knowledge of this behavior is used by Bonnier and Hontebeyrie[47] to design a better fitting procedure. They used the 14 term high temperature series of Luscher and Weisz[48] to obtain an expansion for the susceptibility. The result is a good fit to Monte Carlo data of Berg and Luscher[9] for $\beta \leq 1.7$, providing a better fit to the deviation from asymptotic scaling.

Recently, using his cluster algorithm, Wolff[13] has investigated the O(3) model up to $\beta = 1.9$, where the correlation length is 121. He finds that, even at this large β, asymptotic scaling does not hold for the SA. Finally, Hasenfratz and Niedermayer[19] using different MCRG methods in the region $1.9 \leq \beta \leq 2.26$ see agreement in the discrete β function $\Delta\beta$ with the 2-loop results at $\beta = 2.26$. They also show that asymptotic scaling holds for the TIA starting at a correlation length of ≈ 40 with $m/\Lambda_{\overline{MS}} = 3.4(1)$, and the value for $m/\Lambda_{\overline{MS}}$ agrees with that for the SA at $\beta = 2.26$, $m/\Lambda_{\overline{MS}} = 3.3(1)$.

However the exact value for the mass gap of the O(3) nonlinear σ-model. of Hasenfratz *et al.* [18, 49] is $m/\Lambda_{\overline{MS}} = 8/e \simeq 2.943$. Thus there exists a significant gap between this prediction and the Monte Carlo results. In an attempt to bridge the gap between these results Wolff[22] compared his cluster Monte Carlo results[13] with the exact result using a redefined inverse temperature β_E, and saw behavior much closer to asymptotic scaling.

Another non-perturbative method that has been extensively developed recently is the 1/N expansion by Flyvbjerg and collaborators[50]. The three leading orders of the expansion have been obtained and evaluated for small lattices[51]. The comparison to the above Monte Carlo results shows differences that are comparable or smaller than the estimated size of neglected terms. These results have been extended to infinite lattices by finite size scaling[20]

giving behavior that tends slowly towards the exact value for $\xi > 170$ from about 10% away.

Our results are in good agreement with Wolff[13] and Hasenfratz and Niedermayer[19]. In fact they provide one of only two known computationally feasible methods by which the cluster algorithm results can be confirmed (in this region of large correlation lengths). The other method is multigrid Monte Carlo[52], which has used for $\beta \leq 1.7$ We see that there is no dramatic change in the behavior of the mass gap and susceptibility in the range $1.9 < \beta \leq 2.05$. We also see that the behavior in the rescaled temperature β_E remains closer to the exact value, but the agreement does not get better with increasing correlation length. Finally we show that overrelaxation can be used in a manner in which the effective dynamical critical exponent is close to 1.

4.2 The Model

We study the simplest possible Hamiltonian, with only nearest neighbor interactions. The Hamiltonian, which is also called the Standard Action because of the close correspondence to the one space and one time dimension

nonlinear sigma model, is in this case

$$H = -\sum_{<i,j>} \vec{s_i} \cdot \vec{s_j} \tag{4.1}$$

and thus the partition function becomes

$$Z = \int \prod_i d\mu(\vec{s_i}) e^{-\beta H} \tag{4.2}$$

where $\vec{s_i}$ are unit 3-d vectors and $<\ ,\ >$ denotes the inclusion of nearest neighbor sites only in the sum. Here $d\mu(\vec{s})$ is the measure on the sphere and $\beta = 1/T$ is the inverse of the dimensionless temperature T.

Its continuum limit is the O(3) nonlinear sigma model in one space and one (imaginary) time dimension. Of interest is the behavior of this model in the low temperature, weak coupling limit. This was shown[1] to be a continuation of the high temperature phase of exponential correlation functions. The model is asymptotically free, and a renormalization group calculation has given the beta function of the theory to two-loops[4]:

$$\beta(T) = -\frac{dT}{d(lna)} = -\frac{1}{2\pi}T^2 - \frac{1}{(2\pi)^2}T^3 + O(T^4) \tag{4.3}$$

where a is the lattice spacing. These terms are universal - *i.e.*, the same for

any regularization. Using this and the anomalous dimension we obtain the mass gap m and susceptibility χ of the theory:

$$m = C(2\pi\beta)\exp(-2\pi\beta)\ \{1 + a_1/\beta + O(1/\beta^2)\} \tag{4.4}$$

$$\chi = C'\beta^{-4}\ \exp(4\pi\beta)\ \left\{1 + b_1/\beta + O(1/\beta^2)\right\} \tag{4.5}$$

where C and C' are constants that cannot be calculated in perturbation theory.

The terms with coefficients a_1 and b_1 are the first non-universal terms and arise for three-loop diagrams. Falcioni and Treves[6] calculated these by computing of the third-loop contribution to the beta and gamma functions for the standard action. Their values are $a_1 = 0.575/2\pi$ and $b_1 = 0.0$; these have been confirmed by independent calculation[53].

4.3 The Simulation.

To produce the sample configurations of our Monte Carlo simulation we use a hybrid of microcanonical overrelaxation[44, 54, 55], and the Metropolis algorithm. For an O(N) spin-model, the rule used to obtain a new spin by microcanonical overrelaxation (μOR) is to reflect the old spin $\vec{s}_{old}$ through

the direction of the sum $\vec{\Sigma}$ of its neighboring spins. If $\hat{\Sigma} = |\vec{\Sigma}|^{-1}\vec{\Sigma}$, then

$$\vec{s}_{new} = 2(\hat{\Sigma} \cdot \vec{s}_{old})\hat{\Sigma} - \vec{s}_{old} \tag{4.6}$$

This provides the largest possible step while preserving the energy. To provide ergodicity OR updates are interleaved with Metropolis updates. In a simulation of the XY model[25] this combination drastically improved critical slowing down, giving $z = 1.48$ and 1.2 for number of overrelaxation steps $N_{or} = 8$ and 15 respectively. The number of Metropolis steps was kept constant at $N_{met} = 2$.

In each Metropolis step we construct a new trial spin by adding to the old spin a random vector of fixed length α. The resulting vector is normalized and then accepted or rejected using the usual Metropolis criterion[42]. α is an adjustable parameter, chosen so as to give acceptances between 50 and 55 per cent. The random vectors are constructed to sample a uniform distribution on an S^2 sphere.

A 'sweep' is made up of a number of overrelaxation sweeps (N_{or}) , and a number of Metropolis sweeps (N_{met}). Measurements are made every 'sweep'. The errors in all quantities except the correlation length have been computed by binning the data in groups of 500 for 256^2 lattices, 200 for 512^2, and 100

Beta	L,T	#or(K)	Eff.	N_{or}	Energy	χ	ξ
1.50	256	12600	18	8,12	1.20324(7)	176.4 ± 0.2	11.05(1)
1.60	256	12400	38	12,15	1.27141(7)	448.4 ± 0.7	19.00(2)
1.70	256	9100	85	12	1.32843(7)	1263.4 ± 2.9	34.39(6)
1.70	512	11400	85	20,30,35	1.32848(4)	1263.7 ± 3.3	34.44(6)
1.75	512	7700	66	40,60,120	1.35329(5)	2208.1 ± 6.8	47.4(2)
1.75	768	8000	98	50	1.35322(6)	2197 ± 15	47.2(2)
1.80	512	10100	143	40,45	1.37599(4)	3845 ± 11	64.7(3)
1.80	768	3100	141	40	1.37587(6)	3823 ± 21	64.5(5)
1.85	768	11200	185	60,80	1.39667(3)	6732 ± 25	88.7(5)
1.90	1024,512	6000	184		1.41583(4)	11602 ± 59	121.5(1.1)
1.90	1024	5900	263	100,120	1.41582(2)	11867 ± 62	122.7(1.1)
1.95	1024	700	330	200	1.43363(10)	20640 ± 310	164.8(5.3)
2.00	1024	2700	420	250,300	1.45022(6)	35100 ± 400	224.3(4.2)
2.05	1024	1800	510	300	1.46578(7)	56220 ± 550	295.6(5.2)

Table 4.1: Data from Monte Carlo runs: β, lattice length in each direction, total number of OR sweeps in thousands, average effort (*i.e.*, decorrelation time in terms of OR sweeps), number of OR steps per full sweep, energy per lattice site, susceptibility and correlation length.

for larger. We also calculate the 'sweep' to 'sweep' correlation by measuring the autocorrelation time of the magnetization $\tau_{\vec{M}} \equiv \tau$. Table 4.1 summarizes our results.

4.4 Results

A comparison with Wolff[13] of the energy per site $E = < \vec{s}_i \cdot (\vec{s}_{i+\hat{x}} + \vec{s}_{i+\hat{y}}) >$, susceptibility χ and correlation length ξ shows good agreement on nearly all the values. There are two estimates which disagree to any significant degree,

χ at $\beta = 1.6$ by 2.6σ and the energy at $\beta = 1.7$ by 2.8σ. Our value for the latter is in good agreement with the result of a simulation using MultiGrid Monte Carlo[52] . Our results are consistent between a number of independent runs in each case, 3 and 2 respectively . The agreement to the high accuracy of these results, e.g., $\frac{\delta\chi}{\chi}$ about 0.3% for most points, provides confirmation that all these new methods work.

4.4.1 Mass-gap

To obtain the correlation length ξ or mass-gap $m = 1/\xi$ we fit the zero-momentum correlation function (CF)[10] to $A \cdot (e^{-mx} + e^{-m(L-x)})$. A fit is done in the interval ξ to 3ξ, where ξ is determined self-consistently. To estimate the statistical error of ξ we split the data into 10 parts and averaged the values obtained from the individual fits. For $\beta \geq 1.9$ we have used the jackknife method [56] to obtain estimates of the statistical error, because the increasing autocorrelation times make the statistics gathered less significant. The two error estimates agree for $\beta = 1.9$. (In Table 4.1 we quote these values.)

To check the stability and significance of these fits, further fits were also done in the intervals $\frac{1}{2}\xi$ to $\frac{3}{2}\xi$, ξ to 2ξ, $\frac{3}{2}\xi$ to $\frac{5}{2}\xi$, up to as large a distance as a fit as can be obtained. In all cases we saw that the estimates obtained for all

subintervals were consistent. We note that the statistical errors increase in this progression. This was expected since the relative error of the correlation function increases with distance - because the variance of the measurements of the CF is roughly constant, while the CF itself falls exponentially. Also for $\beta < 1.9$ the value of the effective mass $\log(C(n)/C(n+1))$ was plotted and in all cases showed a plateau at least between 10 and 3ξ. This allows us to extract a mass-gap with confidence that finite size effects are not significant for $\beta \leq 1.9$.

For our largest values of β the finite lattice size will affect our measurements. To obtain an estimate of the effect on the mass-gap we used the estimate of Luscher[57] of the leading correction to the infinite-volume limit. This can be understood as the self interaction of a particle around the finite box and has been computed using the proposed S matrix.

The correction factor $\theta(\zeta)$ is defined from the ratio of masses measured on a lattice of finite length L to that on an infinite lattice:

$$m(L)/m(\infty) = 1 + \theta(\zeta) \tag{4.7}$$

where $\zeta = L \cdot m(\infty)$. The correction is dominated for large ζ by

$$\theta_0(\zeta) = 8\pi \int_0^\infty \frac{1}{t^2 + \frac{4}{9}\pi^2} e^{-\zeta \cosh(t)} dt \tag{4.8}$$

with $\theta(\zeta) = \theta_0(\zeta) + O(e^{-\alpha\zeta})$ where $\alpha > \sqrt{3/2}$. For large ζ, $\theta(\zeta)$ obviously behaves like $e^{-\zeta}$. These result were tested and used extensively in conjunction with Monte Carlo results in section 4 of Bender *et al.* [58], showing that for $z = m(L)L > 1.5$ a very good estimate of the correction is

$$\theta_0(z \cdot [1 - \theta_0(z)]) \tag{4.9}$$

For all our runs with $\beta \leq 1.95$ except that at $\beta = 1.90$ with $L = 512$ and $T = 1024$, the ratio of the lattice length and the correlation length, $z = \frac{L}{\xi}$, is larger than 6. Thus the correction to the correlation length is very small, *i.e.*, $\theta_0(\zeta) < 2 \cdot 10^{-3}$ Only for the cases $\beta \geq 2.0$ and the one singled out above is the correction appreciable. With $3.46 \leq z \leq 4.56$ we use equation (4.9) to obtain an estimate of the correction $\theta(z)$. This is a valid estimate only if the length in one direction is very much larger than the correlation length, but the numerical results of Bender *et al.* lead us to believe that for all these cases they provide a good estimate of the correction. For $L = 1024$, the correction is only appreciable for $\beta = 2.05$, where $\theta_0(z) = 2.9\%$; even for $\beta = 2.00$ it is

only $\theta_0(z) = 0.77\%$. The corrected values are $\xi(\beta = 2.00) = 226.0 \pm 4.2$ and $\xi(\beta = 2.05) = 304.1 \pm 5.3$.

To compare the behavior of the correlation length with asymptotic scaling predictions we use the correlation length defect δ_ξ. This is obtained by dividing the correlation length by the 2-loop result, *i.e.*, equation (4.4):

$$\delta_\xi = \beta\, e^{-2\pi\beta}\, \xi \tag{4.10}$$

Obviously asymptotic scaling is seen if δ_ξ goes to a constant as $\beta \to \infty$. Figure 4.3 shows that asymptotic scaling does not set in for $\beta < 2.00$, but it is not possible to draw a clear conclusion for $\beta \geq 2.00$. We note that the finite size correction affected our evaluation of δ_ξ, so we must expect it to be important in the case of the susceptibility.

In order to compare with the exact result of Hasenfratz *et al.* [18] we use our estimates for the correlation length to calculate the ratio of the mass-gap to the Λ parameter $m/\Lambda_{\overline{MS}}$. In our case it is given by

$$m/\Lambda_{\overline{MS}} = [2\pi\delta_\xi\, (\Lambda_L/\Lambda_{\overline{MS}})]^{-1} \tag{4.11}$$

where $\Lambda_{\overline{MS}}/\Lambda_L = 27.31$ is the ratio of the lattice and modified minimal subtraction scheme Λ parameters calculated by Parisi[7].As figure 4.2 shows,

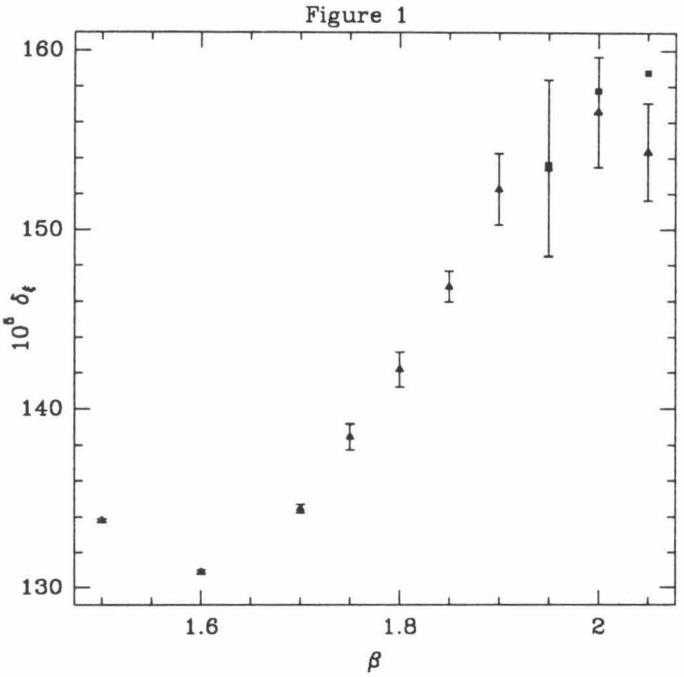

Figure 4.1: Correlation length defect δ_ξ , *i.e.*, correlation length scaled by the 2-loop form. The solid squares are corrected for the expected finite size effect. Note that the errors on these are the same as those for the uncorrected points.

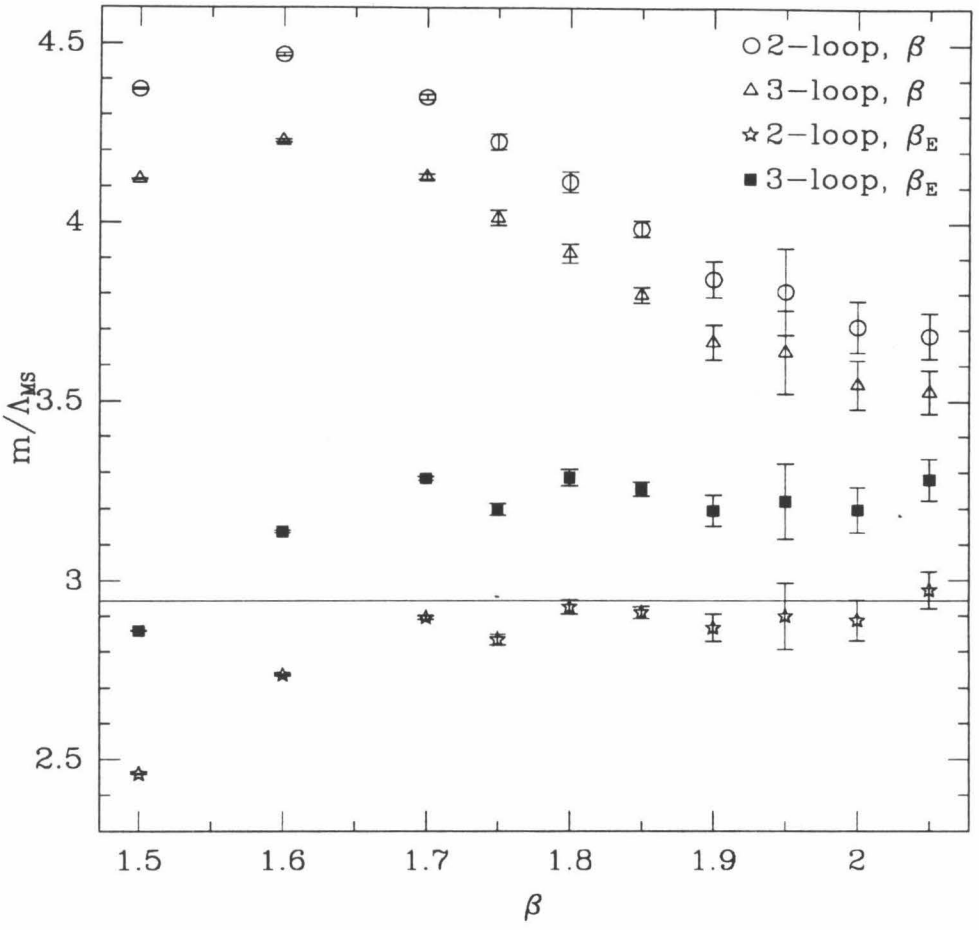

Figure 4.2: The estimates of the ratio $m/\Lambda_{\overline{MS}}$ to 2- and 3-loops in terms of β and $\beta_E = 1/(2-E)$ vs. the inverse temperature β.

using the 3-loop correction term moves our results towards the analytical result, but by a small amount compared to the distance from it. Thus at $\beta = 2.05$ our estimate $m/\Lambda_{\overline{MS}} = 3.52(6)$ is 20% higher than the analytical result.

We can also compare with the results of Hasenfratz and Niedermayer[19] who used Monte Carlo Renormalization Group (MCRG) methods to obtain estimates of the discrete beta function. The values are consistent with asymptotic scaling for $\beta \geq 2.14$, giving $m/\Lambda_{\overline{MS}} = 3.35(9)$. To compare directly with our results we used their data and the 3-loop correction to obtain $m/\Lambda_{\overline{MS}}(\beta = 2.02) = 3.47(8)$. An interpolation of our results at the two neighboring points, yields $m/\Lambda_{\overline{MS}}(\beta = 2.02) = 3.55(5)$; we see that the two estimates differ by about one standard deviation. Our results are thus consistent with those of the MCRG calculation, although they tend to favor a slower fall of $m/\Lambda_{\overline{MS}}$ towards the exact result.

A comparison to the results of the third order 1/N expansion (see the figure in H. Flyvbjerg *et al.* [20]), shows that for increasing β their estimate of $m/\Lambda_{\overline{MS}}$ and our measurements are tending closer.

Another approach to the problem of asymptotic scaling, proposed by S.Samuel *et al.* [21] and tried recently by Wolff[22], uses a redefined inverse temperature derived from the energy. This is an alternative bare coupling,

and effectively performs an infinite order resummation; there are arguments that asymptotic scaling in β_E should be observed earlier[21]. β_E is defined from the $O(1/\beta)$ perturbation expansion of the energy . For our definition of the energy ($E_{ours} = 2E_{Wolff}$) it is $\beta_E = (2 - E)^{-1}$. As figure 4.2 shows, our results agree with Wolff that asymptotic scaling in β_E (*i.e.*, the 2-loop curve) is better and much closer to the exact value predicted by Hasenfratz *et al.* [18]. However the effect of the 3rd-loop correction (for which we calculated that $a_1 = 0.575/(2\pi) + 1/8 - 499\pi/4000$ - see appendix A) moves the result away from the analytical result and towards the results scaled by β. We note that the statistical error of our estimates of the energy are much smaller than those of Wolff[13], e.g., for $\beta > 1.6$ it is more than an order of magnitude smaller. This makes the contribution of the statistical error of β_E to the error of our estimates of $m/\Lambda_{\overline{MS}}(\beta_E)$ negligible, which is not the case for the points at higher values of β of Wolff[22].

4.4.2 Susceptibility

We measured the susceptibility on lattices of different size for $\beta = 1.70$ to 1.80 and at 1.90. The agreement, within errors, seen for $1.70 \leq \beta \leq 1.80$ shows that the finite size effects are very small, so that, effectively, the infinite volume limit has been reached. The disagreement at $\beta = 1.90$ shows that

there is a finite size effect for this point on the smaller lattice, but the data for the other βs lead us to believe that the larger lattice gives us an estimate with very small finite size error. We would expect that for $\beta \geq 2.0$ the finite size effect would be significant ; a rough estimate would be something of the same order as that for the mass gap, *i.e.*, a fraction of a percentage point and a few percent for $\beta = 2.00$ and $\beta = 2.05$ respectively.

To compare with the expected behavior, from equation 5, we divide this behavior out and get a 'scaled susceptibility' or susceptibility defect δ_χ Berg and Luscher[9] :

$$\delta_\chi = 2 \cdot 10^5 \ \beta^4 \ e^{-4\pi\beta} \ \chi \tag{4.12}$$

This should behave as a power series in $T = \frac{1}{\beta}$, and approach a constant for $T \rightarrow 0$. From figure 4.3 it is obvious that we have not reached the region of β where a constant can be extracted, and that the susceptibility for $\beta = 2.05$ at least suffers from finite size effects. We note that for the standard action the third-loop term for the susceptibility is 0 to the accuracy calculated[6], and thus doesn't affect this result.

We also compare the behavior of the susceptibility with predictions based on the assumption of complex singularities in $\chi(\beta)$. Table 4.2 shows another susceptibility defect G_3 for our data and the values obtained by a sophisticated Pade approximant of the 14 term high temperature series. This shows

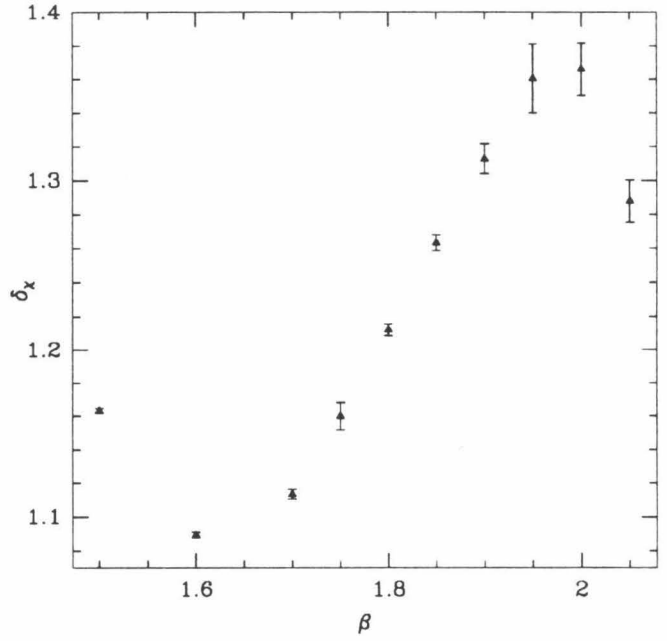

Figure 4.3: Susceptibility defect δ_χ , *i.e.*, susceptibility scaled by the2-loop form. Note that the 3-loop correction is zero.

β	G_3		
	MC	Pade	BL
1.50	1028.5 ± 0.3	1015	1048
1.60	1011.8 ± 0.4	983	962
1.70	1017.4 ± 0.7	964	927
1.75	1027.8 ± 1.8		
1.80	1039.2 ± 0.8		
1.85	1050.0 ± 1.0		
1.90	1055.0 ± 1.7		
1.95	1069.6 ± 4.0		
2.00	1070.7 ± 3.1		
2.05	1055.0 ± 2.6		
∞		944	

Table 4.2: Comparison of results for another susceptibility scaling defect $G_3 \propto \delta_\chi^{1/4}$. Our Monte Carlo estimates are labeled MC, the Pade Approximant of Bonnier and Hontebeyrie[47] is Pade and the numbers of Berg and Luscher[9] are BL.

that this approximant is unable to adequately describe the defect in this region of β, although it does a good job of coarsely describing the significant deviations from asymptotic scaling at smaller values of β.

It would be interesting to extend this work by incorporating the high precision numerical estimates into the Pade approximants to obtain a better fit for the behavior of the susceptibility. This would require the use of multipoint Pade approximants. Also a set of values for χ at different beta values would be used and the results averaged, in order to simulate the statistical errors in the Monte Carlo estimates of χ.

4.4.3 Dynamical critical exponent of overrelaxation

The decorrelation time τ is used to measure the speed with which new, *i.e.*, statistically independent, configurations are generated. Its dependence on the correlation length is parameterized by the dynamical critical exponent z :

$$\tau = c \cdot \xi^z \tag{4.13}$$

Most local algorithms, like Metropolis and heat-bath, have $z \geq 2$. For a free field overrelaxation[24] gives $z = 1$. Neuberger[59] argues that for an interacting field, z should not change substantially from this. Our previous work with the same algorithm as we use for the O(2) or XY model[25] measured τ_{exp} (defined, e.g., by Sokal[60]) and gave $z = 1.48$ for $N_{or} = 8$ and $z = 1.2$ for $N_{or} = 15$ where $N_{met} = 2$. For the $O(4)$ model Heller and Neuberger[61] used another variation of overrelaxation and showed that in 1 dimension $z = 1$, but could not determine it for 2 dimensions.

We obtain the decorrelation time by measuring the auto-correlation function of the magnetization. We will use $c(n)$ to denote this, where n is a distance in number of MC sweeps. For the integrated decorrelation time τ_{int}

and its estimator τ_{int} we use the definitions of Madras and Sokal[62]

$$\tau_{int} = \sum_{n=-\infty}^{\infty} \frac{c(n)}{c(0)} \quad , \qquad \widehat{\tau}_{int} = \sum_{n=-M}^{M} \frac{c(n)}{c(0)} \tag{4.14}$$

with M a multiple of τ, in our case $M = 4\tau$, defined self-consistently. Using these measurements and fitting to the equation 4.13 for $N_{or} = 12$ gives $z = 1.33(1)$.

However we discovered that it is possible to improve on this substantially. We note that when ξ increases and more work is required to produce a decorrelated configuration, it is natural to increase N_{or}. This allows us to perform measurements only on configurations that are less correlated. What we observed was that the performance of the algorithm improves. To compare the speed of decorrelation between runs with different N_{or} we define a new quantity which we call 'effort'

$$e = N_{or} \cdot \tau. \tag{4.15}$$

It is roughly proportional to the computational effort expended to obtain a configuration 1 τ away.

We found that we can define a new exponent z' from $e \sim \xi^{z'}$ when N_{or} is tuned to keep τ constant. This choice was made because we observed that

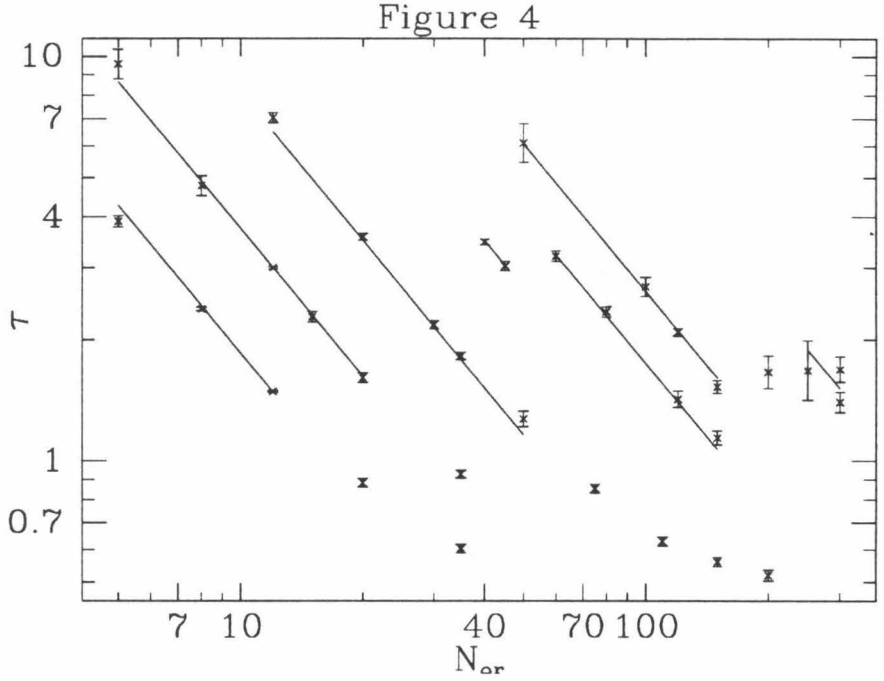

Figure 4.4: Decorrelation time τ vs. number of overrelaxation steps N_{or} for different values of β. Solid lines are the fit to equation (11), *i.e.*, $\tau \propto \xi^z N_{or}^{-z/z'}$.

the effort has a plateau at almost the same value of τ for every ξ. We also found that the behavior of the decorrelation time can be approximated over a good range by

$$\tau = C'' \cdot \xi^z \cdot N_{or}^{-z/z'} \tag{4.16}$$

A fit to $\log \tau = c'' + z \log \xi - \frac{z}{z'} \log N_{or}$ gives $z' = 1.1(1)$, for τ in the ranges $1.1 \rightarrow 1.8$, $2.1 \rightarrow 2.4$ and $3.0 \rightarrow 3.6$. This indicates we have achieved a considerable improvement . A fit to the set of points (N_{or}, ξ , $\tau > 1.0$) gives $z = 1.301 \pm 0.012$, $z' = 1.079 \pm 0.010$ and has a $\chi^2/dof = 1.86$ for 32 degrees of freedom. Figure 4.4 shows the decorrelation time τ vs. the number of overrelaxation sweeps for different values of the coupling β. The solid lines show the fit to the above equation. We note that the points for $\tau < 1.0$ were not included in these fits because they do not follow equation (4.16). Finally, the values for z and z' we measure are, of course, only effective values because in the limited range of τ and ξ we worked in, logarithmic corrections can mask the true (limit) values.

We also try to fit to a general scaling function[63] by plotting τ/ξ^d vs. N_{or}/ξ^f . A plot for $f = 1.0$ and $d = 0.0$ shows that this is close to the correct behavior. Using our knowledge from (11) we constrained the values of d and f to one free parameter with $d = 1.33 - fz/z'$. For $1.06 < f < 1.14$ all points lie on a universal curve, but we tend to favor the central value $f = 1.1$.

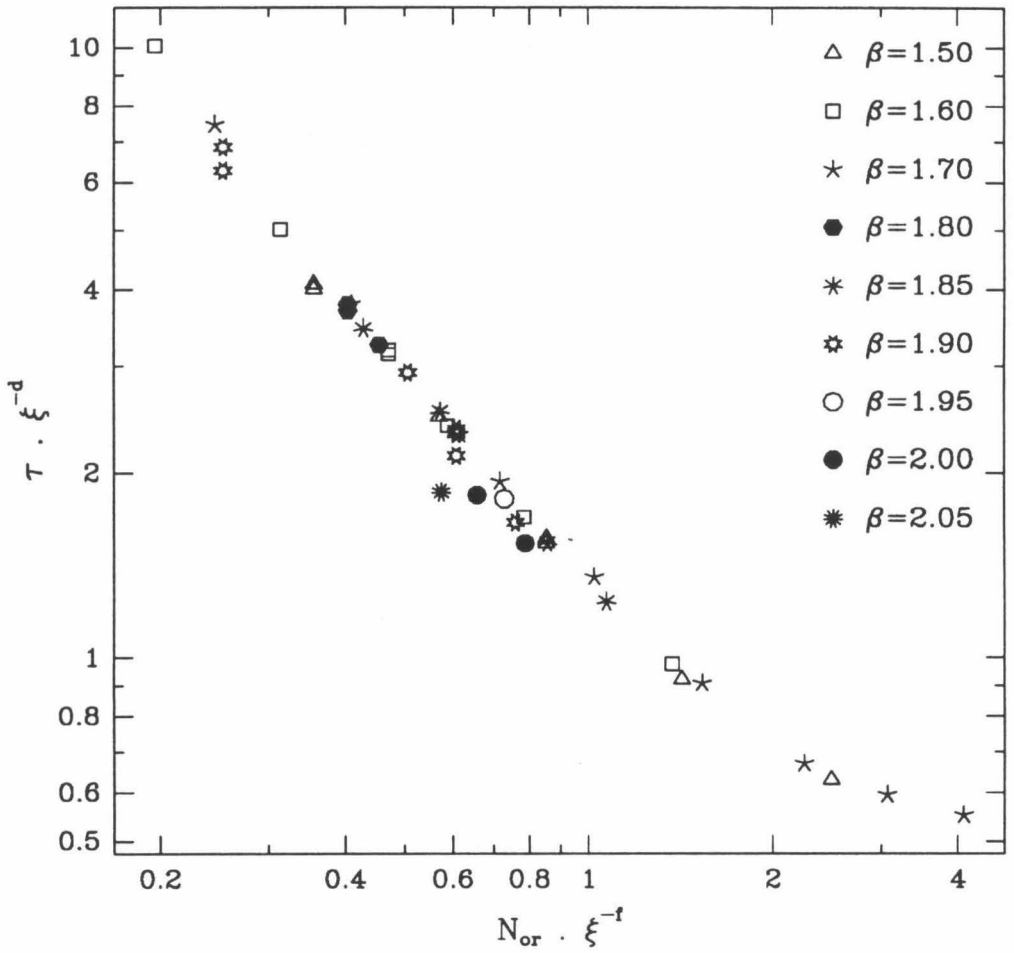

Figure 4.5: Plot for test of scaling of decorrelation time τ, number of overrelaxation steps N_{or} and correlation length ξ according to $\tau/\xi^d = f(N_{or}/\xi^f)$ including data from all the different couplings β. This plot is for the values $f = 1.1$ and $d = -0.017$.

Figure 4.5 shows the data in this case. Only the points for $\beta > 1.95$ do not lie on a universal curve; such a deviation is expected for those points with small values of L/ξ . Another way of presenting the relationship between these quantities is used in figure 4.6. This shows a scaled effort e/ξ^{f+d} (instead of a scaled τ) versus N_{or}/ξ^f for all values of β. The tightness of the points in these plots around a single curve demonstrates conclusively that, for τ roughly constant, the effort $e \propto \xi^{1.1}$.

We can attempt to understand the lower value of z' in the following way: the overrelaxation algorithm has a tendency to decorrelate much faster than other local algorithms, *i.e.*, with an exponent close to the free field value of 1 . The addition of the Metropolis steps destructively interferes with it. The μOR algorithm moves on a deterministic path through phase space. But when ξ is increased the distance in phase space that a set number of overrelaxation steps travels decreases. Thus the addition of Metropolis steps can cause a larger disruption.

This explanation indicates that the effort should flatten out for increasing N_{or}. Our data clearly shows that after a broad plateau the effort slowly increases. This can be seen in figure 4.6, which is a log-log plot. (This effect is also the one that causes the deviation of points with $\tau < 1.0$ from our fit to equation 4.16 .) We can understand the minimum in e vs. N_{or} if we assume

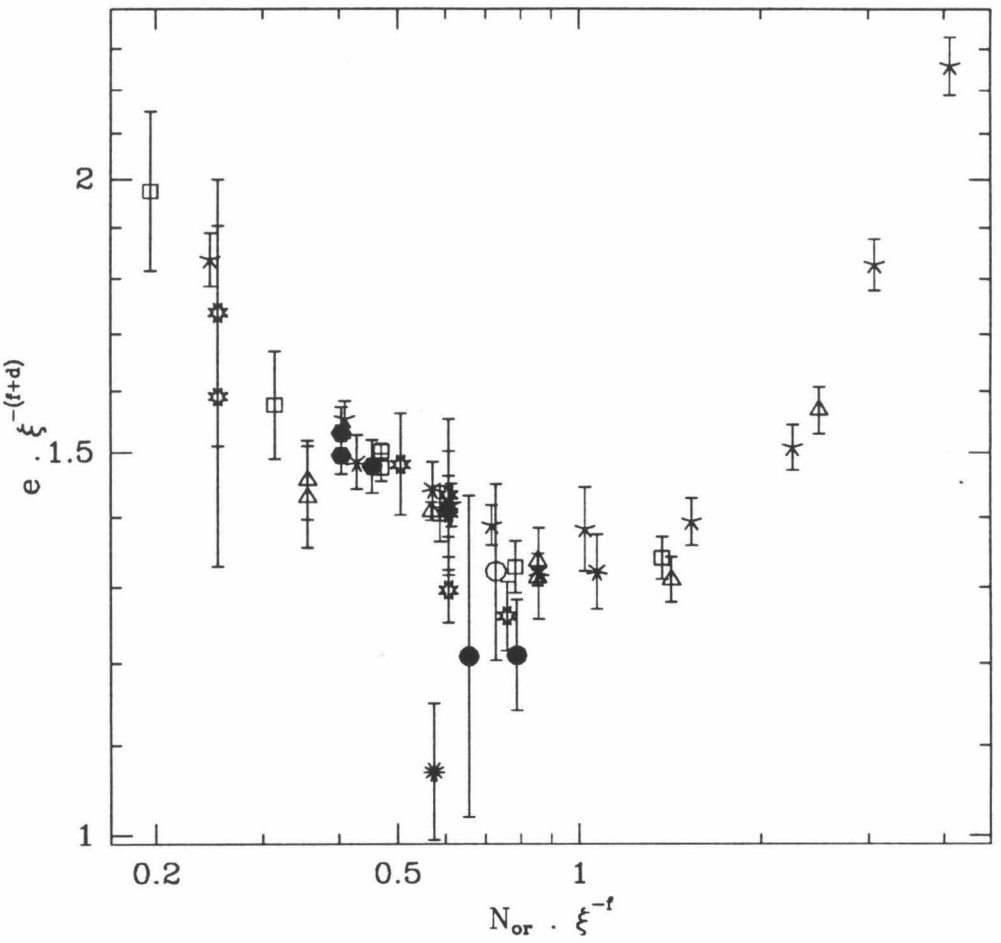

Figure 4.6: Plot similar to fig. 5 for test of scaling of the effort e by e/ξ^{d+f} vs. N_{or}/ξ^{f} , with $f = 1.1$ and $d = -0.017$. The error in N_{or}/ξ^{f} is smaller than the width of the points. The labels are the same as in the previous figure.

that a certain set of (N_{or}, N_{met}) corresponds to a pure, non-microcanonical, overrelaxation algorithm (as used by Heller and Neuberger[61]) with a parameter $\omega \neq 2$. It is obvious that as $\frac{N_{met}}{N_{or}} \rightarrow 0$ that $\omega \rightarrow 2$. What is seen in the case of O(4) in one dimension, by Heller and Neuberger[61] , is that the function $\tau(\omega)$ has a minimum close to $\omega = 2$.

4.5 Conclusions and discussion

Our results confirm those obtained using the Wolff cluster algorithm and extend them to larger lattices and correlation lengths. However asymptotic scaling is not reached with the standard action even at $\beta = 2.05$ and $\xi = 300$ and our results give a value of $m/\Lambda_{\overline{MS}}$ that is 20% higher than the exact result.

We also showed that overrelaxation can be used in a manner in which the effective decorrelation exponent is close to 1, confirming the prediction of Neuberger[59]. This result has been confirmed in an analytical study of a similar algorithm for the Gaussian model by Wolff[26]. He shows that a choice of the mix of overrelaxation sweeps and heat bath sweeps obtains optimal performance with $z = 1$. In that case it was necessary to vary the number of overrelaxation sweeps randomly with a the mean proportional to

the correlation length; however this is a feature of the free model, due to the lack of interactions that mix different Fourier modes. In the case of the O(N) model, changing this mix does not provide any benefit.

By comparison, cluster algorithms do not exist for most models of interest, or even, e.g., for actions with mixed ferromagnetic anti-ferromagnetic terms for the Ising model; any frustration defeats current methods of obtaining practical algorithms. Where they are applicable they require many orders of magnitude fewer arithmetic operations at large correlation lengths - because they have a smaller dynamical critical exponent. However they require special effort to be used on vector or parallel computers. Overrelaxation, and hybrid overrelaxation in particular, remains a simple, vectorizable and efficient algorithm well suited to vector and parallel machines and competitive for many problems. In particular it is the most efficient algorithm discovered to date for pure gauge lattice gauge theory.

It is not clear what causes this effect. One can speculate that another possibility is that they are connected to instantons, which are a property of this model unique amongst O(N) models. A first calculation of the effect of instantons by Evertz[64], however, does not provide an answer. The correction to the beta function calculated has the wrong sign to explain the deviations from asymptotic scaling observed by us. However other O(N) models show

similar but smaller deviations from asymptotic scaling, as seen for the O(4) and O(8) models by Wolff[22]. Since they do not possess instantons this would tend to discount the possibility that effect is due to this topological factor.

The most likely cause are singularities in the partition function for the standard action. These are seen for the susceptibility in the complex β plane by Butera *et al.* [23]. The same effect is also seen in models interpolating between the XY and O(3) models using a mass term to suppress one component of the O(3) field[65] . A critical line in the coupling constant space extends from the critical point of the XY model towards the region of non-monotonic behavior of the beta function we saw in the O(3) model

Additional simulations

After the above work was published, additional simulations were done using a cluster algorithm. We simulated at the same couplings $2.00 \leq \beta \leq 2.05$ on a 2048^2 lattice and also extended to $\beta \leq 2.15$ on the same size lattice. The results supported our previous conclusions. The Monte Carlo update algorithm we used was the Wolff Wolff embedding[29, 60] of the $O(N)$ spins onto Ising spins, which were simulated by Swendsen-Wang dynamics. The same method is used for the update in the MCRG study in chapter 6. Here we wish to mention some of the results, which will later be referred to

for comparisons in chapter 6.

First, we verified the implementation of our algorithm by measuring the energy, susceptibility and correlation length at $\beta = 1.5$ on a 256^2 lattice and at $\beta = 1.9$ on a 1024^2 lattice. For the susceptibility in addition to the usual estimator $\chi = \frac{1}{V} < \vec{M}^2 >$, where $\vec{M}$ is the magnetization - *i.e.*, the total spin of the lattice, we also used the improved cluster estimator of Wolff[29]. We also used a variant of the correlation function estimator of Wolff[13] to measure the zero momentum correlation function. We quote the improved correlation length (derived from this correlation function) and the improved susceptibility in tables 4.3 and 4.4, in addition to the values from the usual estimators.

Simulations were done on a 2048^2 lattice as this was the largest that could fit on the large memory 16,384 (16K) processor Connection Machine-2 we used. On this lattice size simulations beyond $\beta \geq 2.15$ do not seem to be of value when one is seeking the infinite volume value of mass-gap and susceptibility, since even at $\beta = 2.15$ the correlation length cannot be reliably extracted from the data for the zero-momentum correlation function.

Comparing with the values of the correlation length estimated using the overrelaxation algorithm on 1024^2 lattices, the current estimates are in reasonable agreement. Agreement between improved and standard estimators

β	#	Energy	ξ	$\xi_{improved}$
2.00	13750	1.45023(2)	228.9 ± 6.0	
2.00	13300	1.45022(2)	245.0 ± 12	
2.00	7600	1.45025(3)	221.1 ± 8.0	225.0 ± 3.2
2.05	34000	1.46578(2)	299.7 ± 5.5	
2.05	19500	1.46578(2)	315.5 ± 7.5	312.3 ± 4.2
2.10	28200	1.48036(2)	412.2 ± 9.3	415.6 ± 5.1
2.10	37200	1.48035(1)	426.3 ± 7.1	

Table 4.3: Data from cluster Monte Carlo simulations at $\beta = 2.0$ to 2.1 on a 2048^2 lattice: Energy and correlation length estimates (the latter obtained from standard and improved zero-momentum correlation function estimators).

β	#	χ	$\chi_{improved}$	τ_{int,χ_i}	$\tau_{int,\vec{M}}$
2.00	13750	$35000 \pm (> 875)$	36970 ± 290	6.26(57)	
2.00	13300	37860 ± 600	36990 ± 250	3.50(24)	2.32 ± 0.14
2.00	7600	35440 ± 700	36492 ± 314	4.02(43)	2.28 ± 0.18
2.05	34000	63060 ± 660	63680 ± 540	5.74(32)	2.44 ± 0.09
2.05	19500	65000 ± 870	64620 ± 470	5.14(37)	2.46 ± 0.12
2.10	28200	106130 ± 1250	108130 ± 820	5.46(33)	2.38 ± 0.10
2.10	37200	110200 ± 1100	109780 ± 740	5.41(28)	2.44 ± 0.08
2.15	4700	176200 ± 4500	174600 ± 3700	7.0(1.1)	2.45 ± 0.24

Table 4.4: Data from cluster Monte Carlo simulations at $\beta = 2.0$ to 2.15 on a 2048^2 lattice. Statistics (#), standard χ and improved $\chi_{impoved}$ estimators for the susceptibility and estimates for the integrated decorrelation time for the latter and the magnetization.

is good. The agreement between different runs for the standard estimator of the susceptibility at $\beta = 2.0$ and 2.1 is not good, but not very bad.

The behavior of the mass gap ration $m/\Lambda_{\overline{MS}}$in this range, to two and three loops, can be seen in the figure 6.9 in the next chapter. The trend towards the exact value continues, but agreement is not seen for $\beta \leq 2.10$.

Chapter 5

New SIMD Algorithms for Cluster Labeling on Parallel Computers

Monte Carlo simulation is a very important numerical technique for studying a wide range of problems in the physical sciences, and in particular, the statistical mechanics of spin models of magnets[66, 67]. Unfortunately, traditional Monte Carlo algorithms for these models, such as the commonly used Metropolis algorithm[42], suffer from *critical slowing down* near the regions of interest – the critical points separating different phases of the system. This means that the autocorrelation time (the number of iterations needed

to generate a new, statistically independent, data point) increases as L^z at the critical point, where L is the linear size of the system, and z is the dynamic critical exponent[68, 60]. z is at least 2 for most local algorithms, such as Metropolis, so the efficiency of these methods decreases rapidly as the size of the system is increased.

The reason for this poor performance is that standard Monte Carlo algorithms are local. In the lattice of spins which represents, for example, a magnetic material, only a single spin at a time is changed, and this change is influenced only by the spins on neighboring sites. Information undergoes a random walk on the lattice, and thus takes a time of order L^2 to propagate throughout the lattice. In the last few years, algorithms have been invented for certain types of spin models which make large-scale, non-local changes, and greatly reduce critical slowing down[28, 29]. In these so-called *cluster* algorithms, clusters of spins (rather than single spins) are changed at each step of the Monte Carlo procedure. The clusters are formed by generating bonds connecting neighboring sites, using a probabilistic procedure which varies between different models and algorithms (for reviews of cluster algorithms, see Refs. 68, 60, 69, 70).

The major computational task of these cluster algorithms is the identification and labeling of the clusters of connected sites, given the configuration

of bonds. This is an instance of a connected component labeling problem for an undirected graph[71, 72], where the vertices are the lattice sites and the edges are the bonds between connected sites. The goal of the component labeling algorithm is to end up with the same label on all connected sites, and different labels for all disconnected clusters.

Sequentially, this can be done in time of order V (the number of vertices, which in our case is the volume, or number of sites, in the lattice), and consequently the cluster algorithms run about as fast as the local algorithms (see Refs. 71, 73, 74 for a discussion of sequential labeling algorithms). However this may not be the case if our computer is a distributed memory parallel machine. Local algorithms perform very efficiently on parallel machines, whereas efficient component labeling on a parallel machine is a very difficult problem[73]. Here the information concerning the connectivity of a given physical part of the lattice is only contained in a single processor, and obtaining information from distant regions of the lattice (and hence also of the computer) can be very slow if the clusters are large, and thus contain sites which are distributed over many processors.

Let us assume that the time taken to label the clusters scales asymptotically as L^{d+y} for a lattice of L^d sites, where y is an exponent indicating *computational slowing down*, in analogy with the dynamical exponent z ex-

pressing the critical slowing down of a Monte Carlo simulation. This means that the overall computational cost of a Monte Carlo cluster algorithm simulation at the critical point will scale as L^{d+y+z}. If we cannot find a parallel labeling algorithm for which y is zero, the advantages of cluster update algorithms over traditional local algorithms may be eliminated on a parallel machine by the computational complexity of labeling the clusters.

Our aim is to find a parallel component labeling algorithm with no computational slowing down (*i.e.*, $y = 0$). We will consider here the case where the parallel computer is a Single Instruction Multiple Data (SIMD) machine, although the ideas described here could also be applied to Multiple Instruction Multiple Data (MIMD) machines. We have implemented all the algorithms on the CM-2 Connection Machine, which is a typical massively parallel SIMD computer[75].

In order to test the algorithms we have studied the clusters formed in the physically interesting case of the Swendsen-Wang cluster algorithm[28] applied to the Ising spin model at its critical point. These clusters are very similar to those created by the simple procedure of randomly connecting neighboring sites on a two-dimensional lattice with probability $\frac{1}{2}$. Clusters created in this way are very difficult to label efficiently on a parallel machine, since the clusters in a particular configuration of the connections come in many

different sizes, have extremely irregular shapes, with small clusters embedded in larger ones, and typically including a very large cluster which will span the lattice. This type of problem is consequently an excellent test of parallel component labeling algorithms.

We also note that the worst case behavior of the labeling algorithms is not relevant for this problem – what we are really interested in is the *average* time to label physically realistic configurations of clusters which occur in the cluster update of the spin model. We have therefore obtained all our data by averaging over a large number (typically 400) of different realizations of the site connections, taken from different Swendsen-Wang bond configurations for the Ising model at its critical point. This is in order to get statistically significant results from which we can obtain the scaling behavior of our algorithms, and timings for our implementations of these algorithms on the CM-2.

5.1 Simple parallel algorithms

The simplest and most obvious SIMD component labeling algorithm is local label propagation[73, 76]. We start with a different label on each site, and with a list of nearest neighbor connections (these will be Boolean variables in the

following: *off* means no connection is present and *on* means that there is a connection). Each site then looks to each of its neighbors in turn. If it is connected to this neighbor, and if its neighbor's label is smaller than its own label, then it replaces its label with that of its neighbor. This procedure is repeated until there is no change to the labels, at which time each cluster will be labeled by the minimum initial label of all the sites in that cluster.

This local algorithm suffers from computational slowing down, and for many problems of interest (such as spin models at their critical point) its performance degrades very fast with increasing volume. This is because there is typically a large cluster whose graph-theoretic diameter or *chemical distance*, defined as the maximum value of the shortest path length between two points in the cluster, scales as L^f, where the exponent f is approximately 1 for the two-dimensional Ising model[77], *i.e.*, the diameter of the largest cluster scales approximately linearly with L. For any local labeling algorithm, the minimum label has to diffuse across this large cluster, so we expect that $y = f \approx 1$.

This algorithm can be improved by making the propagation step non-local. One way of doing this is, instead of propagating the labels only to neighboring connected sites, to propagate them as far as we can along a given direction, until we come to a site with no connection in that direc-

tion. On the CM-2 this can be done very quickly by using the intrinsic **scan_with_minimum** function[78]. This routine operates on a row of numbers (labels in our case), each of which has an associated Boolean flag (the connections). It runs along each connected section of the row and deposits at each site the minimum of the numbers in the section up to that point (this is done in a distance doubling way, taking $\log_2 L$ steps). One step of this labeling method consists of a *scan* in each of the forward and backward directions of every axis. If periodic boundary conditions are used, this must be supplemented by a local label propagation step, since the *scan* routine does not wrap around the lattice.

Another way of improving the above algorithms is the notion of *connection improvement.* So far we have considered the bonds between sites to be static, in other words they are set up at the beginning and remain unchanged throughout the labeling procedure. However it is actually very useful to change, or improve, the connections as the labeling procedure progresses, and we learn more about the connectivity of the sites. It will often happen that neighboring sites will not have a bond between them, but will still be part of the same cluster, as shown in Fig. 5.1. If we compare the labels of neighboring sites at each step of the labeling algorithm, then at some point we will find that these two neighboring sites have the same label.

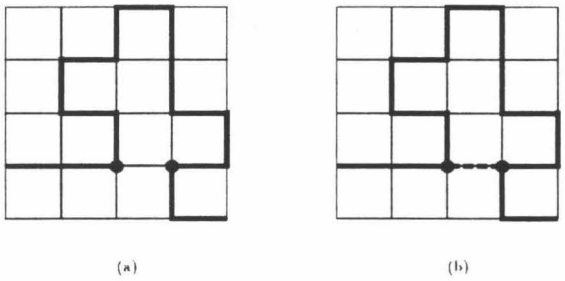

Figure 5.1: An illustration of connection improvement. The original bonds are shown as the thick lines. If at some point in the labeling algorithm it is found that sites i and j (denoted by the filled circles) have the same label, then a new bond (the dashed line) is introduced, so that changes in the labels are now propagated faster between these two points. For the multi-scale algorithm the same idea is used, except that i and j do not have to be neighboring points.

We could then place a connection between these sites, since we now know that they are in the same cluster. Improving the bonds in this way means that new labeling information can now flow directly between these two sites, rather than by an indirect route via the original bonds.

Connection improvement is especially useful when applied to the *scan* algorithm, since in that case the addition of extra connections means that labels may be propagated much further in a single *scan* operation. This can be seen in Fig. 5.2, which shows a log-log plot of the average number of iterations required to complete the cluster labeling for the local algorithm and the *scan* algorithm, both with and without connection improvement. In Table 5.1

algorithm	y
local	1.08(2)
scan	1.09(3)
local improved	1.01(2)
scan improved	0.84(3)

Table 5.1: Exponents y of computational slowing down for some simple component labeling algorithms applied to clusters of Swendsen-Wang bonds for the Ising model.

we show the exponent y for computational slowing down for each of these algorithms, which are obtained from the straight line fits shown in Fig. 5.2. As expected, the exponents are all near 1, except for *scan* with connection improvement, which is substantially smaller, although still far from zero. However we should note that these results are very dependent on the type of bond configurations used. Note for example that configurations for which the clusters are fully connected, smooth, regular shapes, such as may occur in labeling objects in image processing applications, would be labeled in a very small number of *scan* operations. We might expect that y would be zero for the *scan* algorithm for those particular types of configurations.

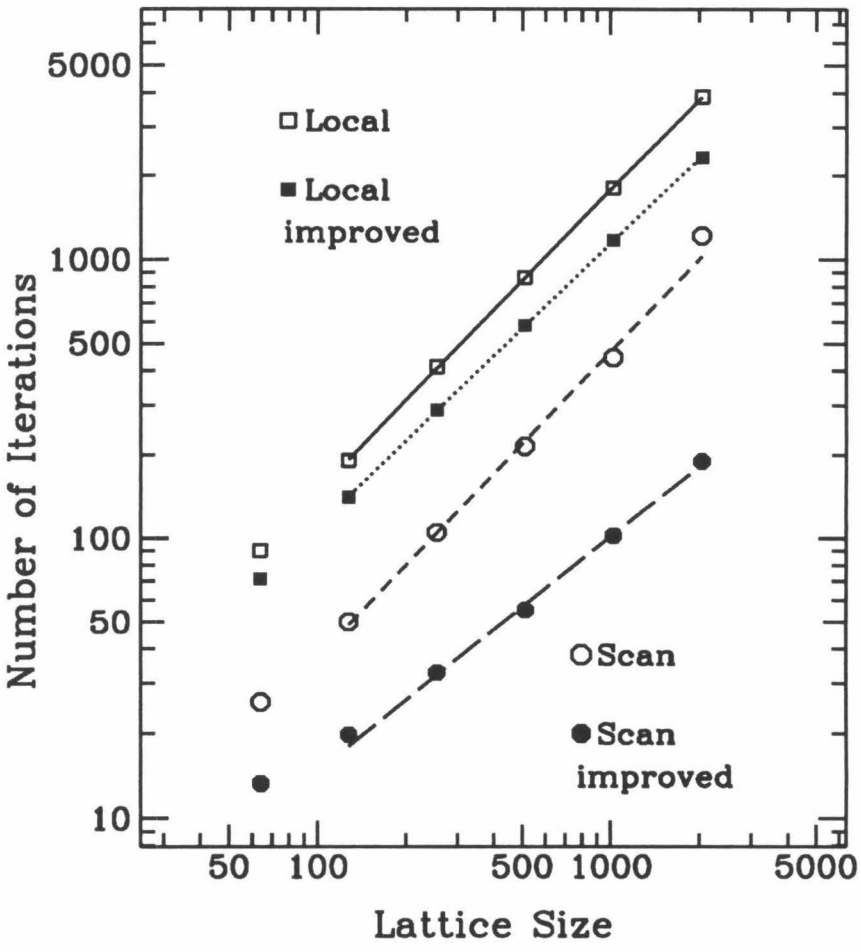

Figure 5.2: A log-log plot of the average number of iterations versus lattice size for labeling Ising model clusters using the local propagation and *scan* algorithms, with and without connection improvement. The errors are not shown, but are smaller than the points.

5.2 A multi-scale algorithm

In this section we describe a regular, synchronous, multi-scale algorithm for cluster labeling[79]. We present numerical evidence that the average number of iterations and the average time taken do not undergo any power-law computational slowing down (*i.e.*, $y = 0$) for our application of labeling Ising model clusters.

The algorithm is effective on a general SIMD machine provided that the switching network has some very basic non-local connections. In the following we will assume that the machine allows very fast communication between sites which are a distance of 2^m sites away in any direction of the physical lattice. These are the only non-local connections we need in order to build an algorithm which is not affected by power-law slowing down. Such connections would be provided, for example, by a machine with a hypercube topology.

On the CM-2 the mapping of the physical structure of the lattice to the (almost) hypercube processor communication network provides specific communication to nodes of the lattice that are at a distance of any power of two away, known as **power_of_two** operations. This involves the transfer of information over not more than two links of the hypercube, and should thus be executed at not less than half the speed of local communications;

however the relative timing compared to a local communication depends on the virtual processor to physical processor (VP) ratio (*i.e.*, the number of lattice sites per processor), as we will see later.

In common with the method proposed by Brower, Tamayo and York[76], this method uses a *multi-scale* approach in propagating cluster labels, in order to overcome the slowing down inherent in local labeling algorithms. However this algorithm is much simpler, and seems to have better scaling properties.

Our algorithm works for a lattice of any dimensionality d, but for ease of description we will consider a two-dimensional problem. In this case the key variables used are Boolean connections that are set up in the x and the y axis at a distance 2^m, for $m = 1, \ldots, l-1$ (where the lattice size $L \equiv 2^l$), by a logical AND of connections at level $m-1$. For example, the distance 2 connection between sites i and $i + 2\hat{x}$ is set (*i.e.*, turned *on*) if both the connections between sites i and $i + \hat{x}$ and sites $i + \hat{x}$ and $i + 2\hat{x}$ are already *on*. These connections are rebuilt in this way at each iteration. In addition to building up the long distance connections in this manner, at each iteration we also use connection improvement, thus a connection between two sites at a generic distance M which was originally *off* can be set (*i.e.*, declared to be *on*) if the two sites are found to have the same label. Using connection

improvement greatly reduces the number of iterations needed to converge to the final values of the labels.

Thus, during one multi-grid label updating cycle each site will look in turn at each of its $2d$ neighbors at each level m of the multi-scale connections. It will update, when possible, its label and also update its connection by merging the level $m-1$ connections and by using connection improvement. A full cycle of the algorithm sweeps all l connection levels, and a single such cycle solves the trivial case where all connections are *on*. As the labeling progresses, what happens is that an increasing fraction of ever longer distance connections are set as sites are recognized as belonging to the same cluster, and these connections become fast long distance communication channels.

In Fig. 5.3 we show the average number of iterations needed to label the Ising clusters as a function of $\log L$. The logarithmic slowing down is very clear. We do not see any sign of power-law behavior, or of a higher power of the logarithm. Each iteration of the algorithm involves a multi-grid cycle of $\log_2 L$ steps, with each step taking approximately the same amount of time, which is proportional to L^d/N, where N is the number of processors ($N \leq L^d$). Thus the total CPU time goes as $L^d(\log L)^2/N$, or $(\log L)^2$ for a machine with L^d processors. Hence this algorithm adds only a $(\log L)^2$ term to the overall slowing down of a spin model cluster algorithm. The average

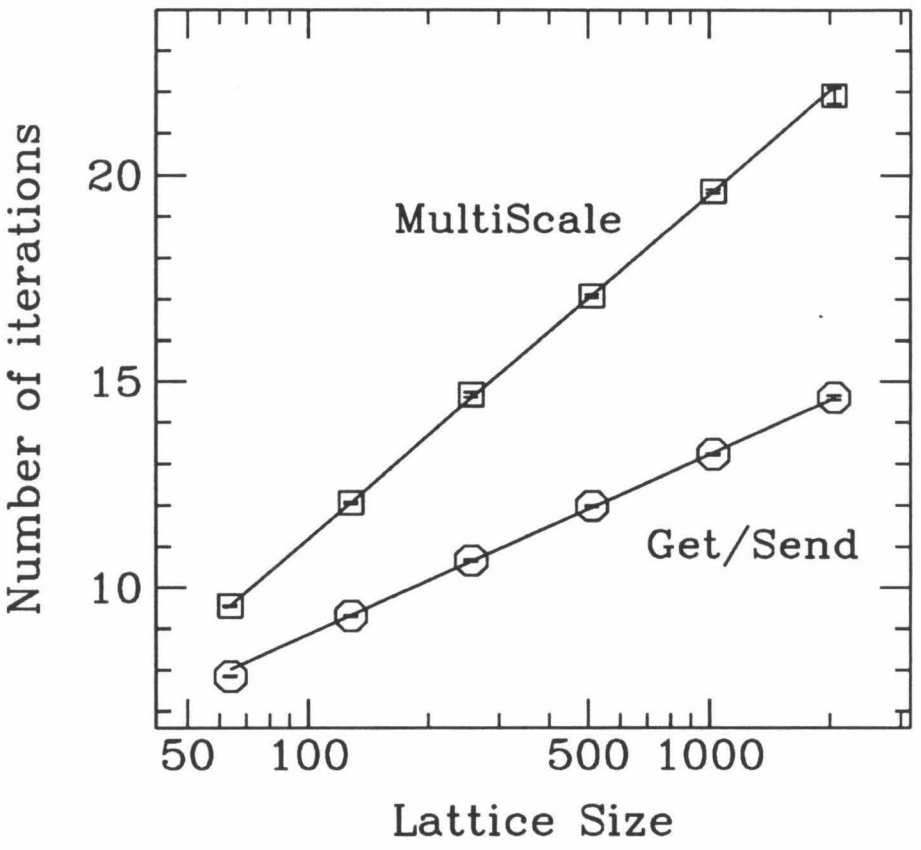

Figure 5.3: Average number of iterations versus log L for labeling Ising model clusters using the full depth multi-scale and the *get/send* algorithms.

labeling time per site as a function of $\log L$ is shown in Fig. 5.4.

The effect of different VP ratios means that the times for the multi-scale algorithm on a fixed number of processors for different lattice sizes do not scale simply as $L^d(\log L)^2/N$. Firstly, we note that local operations are more efficient at higher VP ratios, since a greater proportion of the neighboring sites will be on the same processor, so less inter-processor communication is required. This effect decreases as the depth of the multi-scale procedure is increased, since more sites at a distance 2^m are going to be on different processors as m increases. Eventually the communication distance will be greater than the size of the sub-domain on each processor, so that all data must be communicated between processors. The communication time will therefore be roughly constant at this level and higher, and at the highest levels it is roughly independent of the VP ratio. (The situation is actually slightly more complicated than this, since on the CM-2 there are 16 processors per chip, and it is inter-chip, rather than inter-processor, communication which is costly.) Thus the ratio of the time taken to do a step at the largest depth to the time to do a local labeling step increases from about 2 at $L = 128$ to about 6 at $L = 2048$ on a 16394 (16K) processor CM-2.

There is a way however to combat the higher cost of deep iterations and significantly reduce the running time of our algorithm with only a simple

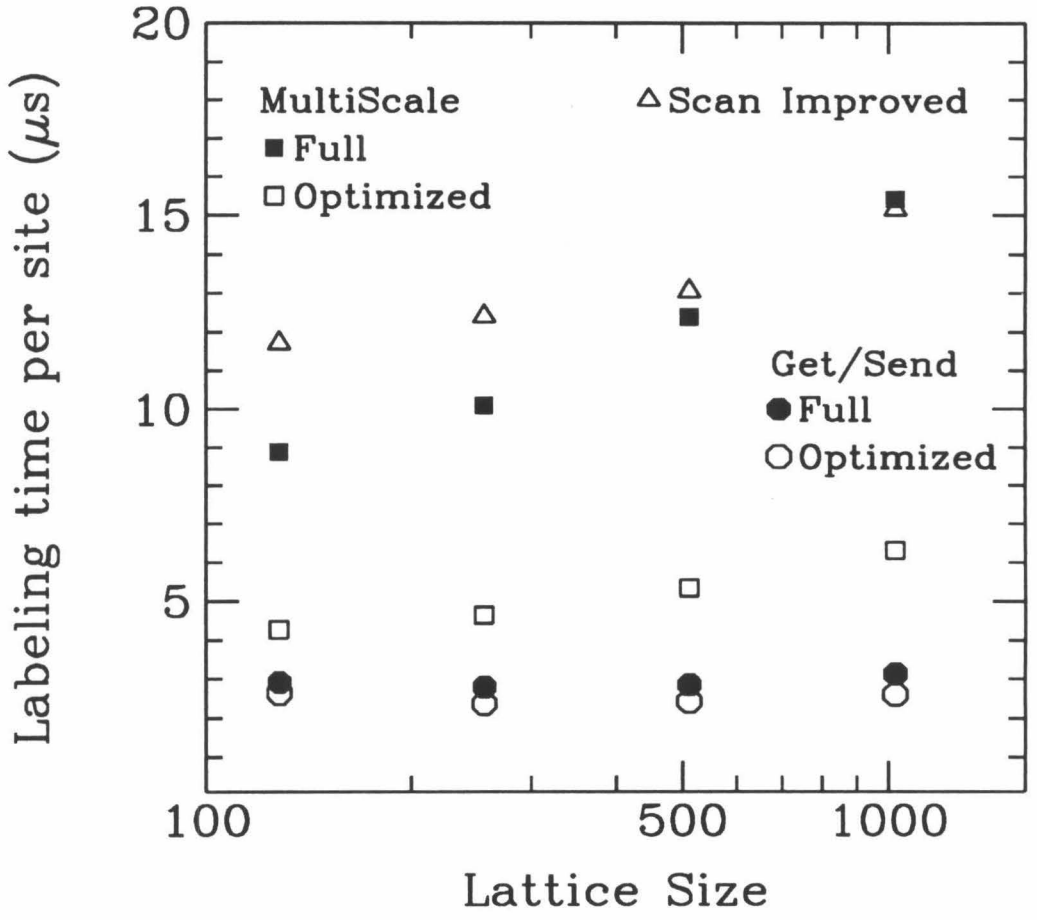

Figure 5.4: Average times per site versus $\log L$ for labeling Ising model clusters on the CM-2 using the *scan* with connection improvement, multi-scale (optimized and full depth) and *get/send* (with and without optimization) algorithms.

modification. Clearly at the beginning of the labeling procedure the long distance connections are all *off*, and due to the fractal structure of the connections, it takes several iterations before a significant number of long distance connections are generated. It is thus very useful to tailor the number of multi-grid levels as a function of the cycle number: a lower depth is useful at the beginning, while using longer distance connections is more useful towards the end of the procedure. Fig. 5.3 shows the average iterations for the simplest case, where the depth is constant, while Fig. 5.4 gives the timings for both the full depth version and the optimized method.

In order to investigate the effect of varying the depth of the multi-grid procedure, we have measured the number of connections at every multi-grid level after each iteration of the full depth algorithm. We show this for a typical configuration in Fig. 5.5. It can be seen that the points where the different levels become useful, *i.e.*, where there are a reasonable number of connections (of the order of 10%, for instance), increases roughly linearly with the number of iterations. We therefore chose in our modified algorithm to make the depth a linear function of the iteration number. Since the usefulness of a connection at a certain level depends on the relative timings of different operations on a specific machine, this relation must be determined empirically.

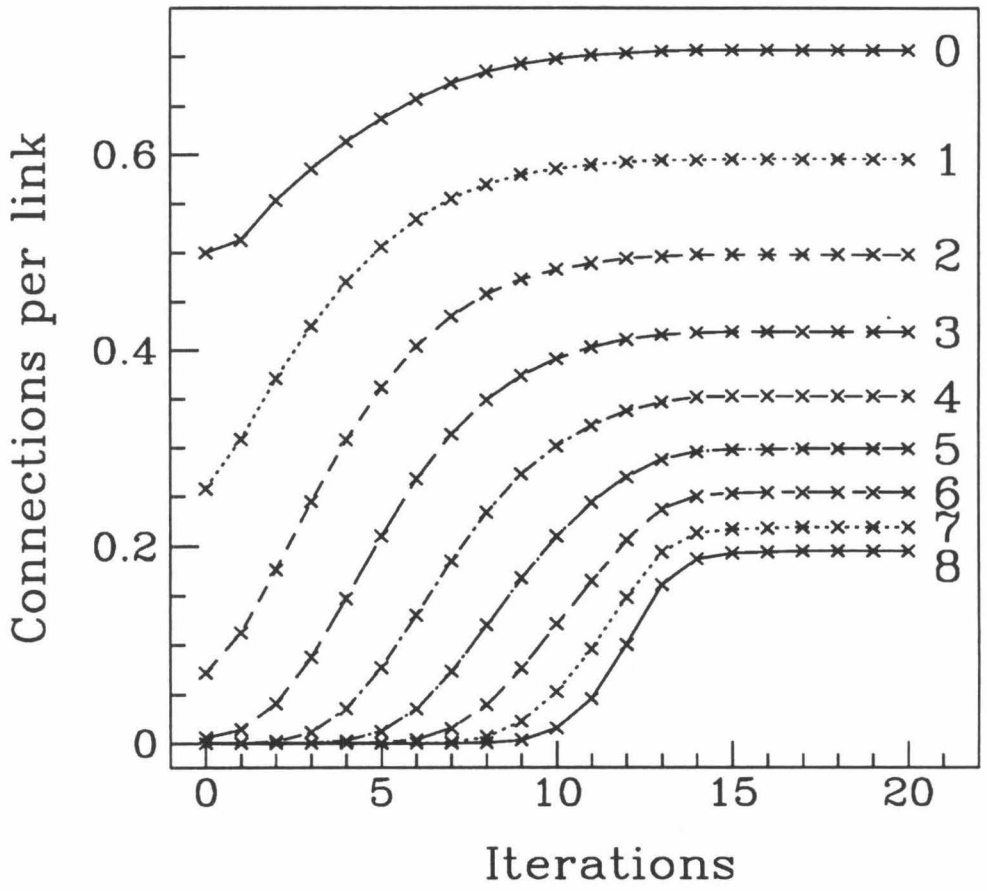

Figure 5.5: The number of connections at different levels (shown at right) as a function of iteration number for the multi-scale algorithm, for a lattice of size $L = 1024$.

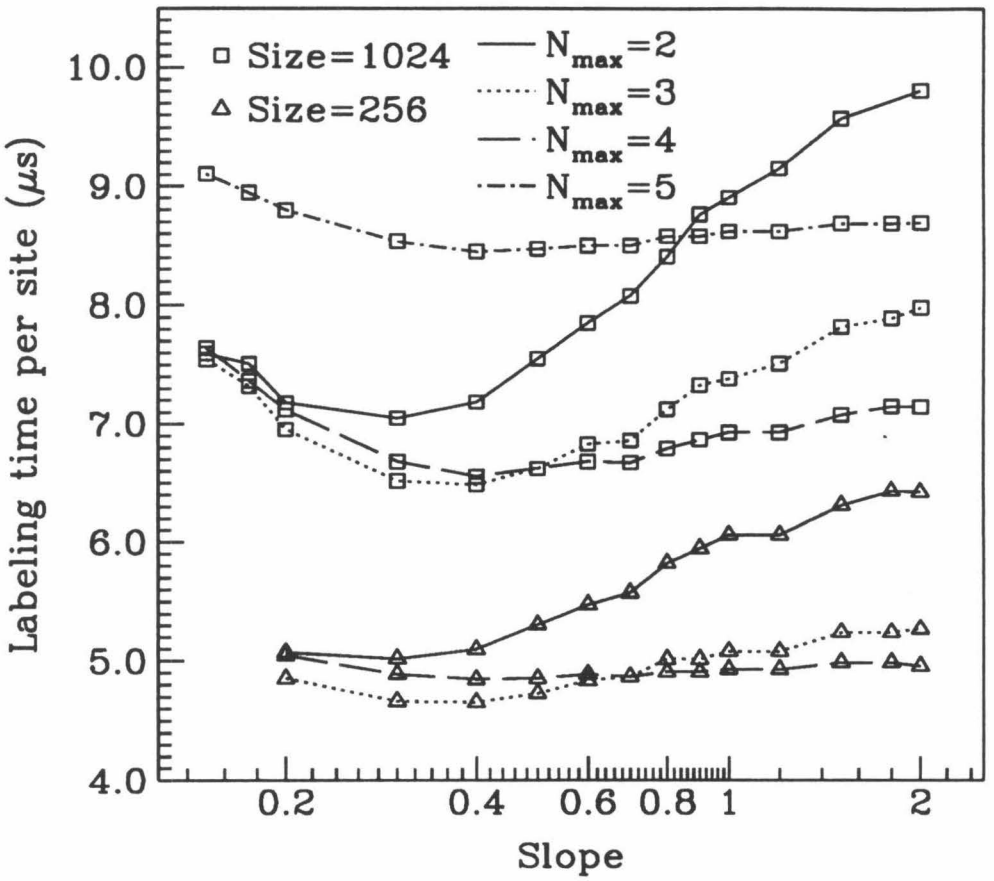

Figure 5.6: Average times per site on the CM-2 for labeling Ising model clusters using the multi-scale algorithm as a function of the slope parameter, for different lattice sizes and maximum depths.

Although Fig. 5.5 shows that a large number of connections at higher levels exist, we found that steps at the highest levels cost too much and were thus of comparatively little use, so a maximum depth of $d_{max} = \log_2 L - N_{max}$ steps was used. We thus parameterize the depth at each iteration by

$$depth = \min\{slope * iteration, d_{max}\}, \tag{5.1}$$

and seek to find the optimum value for the *slope* and the maximum depth parameter N_{max}, which is the number of high level iterations that are not used.

The behavior of the average labeling time versus the slope for some different values of L and N_{max} can be seen in Fig. 5.6. A minimum exists between 0.3 and 0.5 in all cases. The minimum is fairly broad and its breadth tends to increase as L increases. Fig. 5.4 shows the average labeling time per site for the optimized algorithm as well as the full depth multi-scale procedure. Note that the time for the optimized procedure is significantly smaller. The optimal value of N_{max} is 3 (*i.e.*, the three highest levels are not used) for most lattice sizes, although for large L it is slightly more efficient to exclude the fourth highest level as well, since the ratio of the time taken at the higher levels to the time for a local iteration is greater at large L.

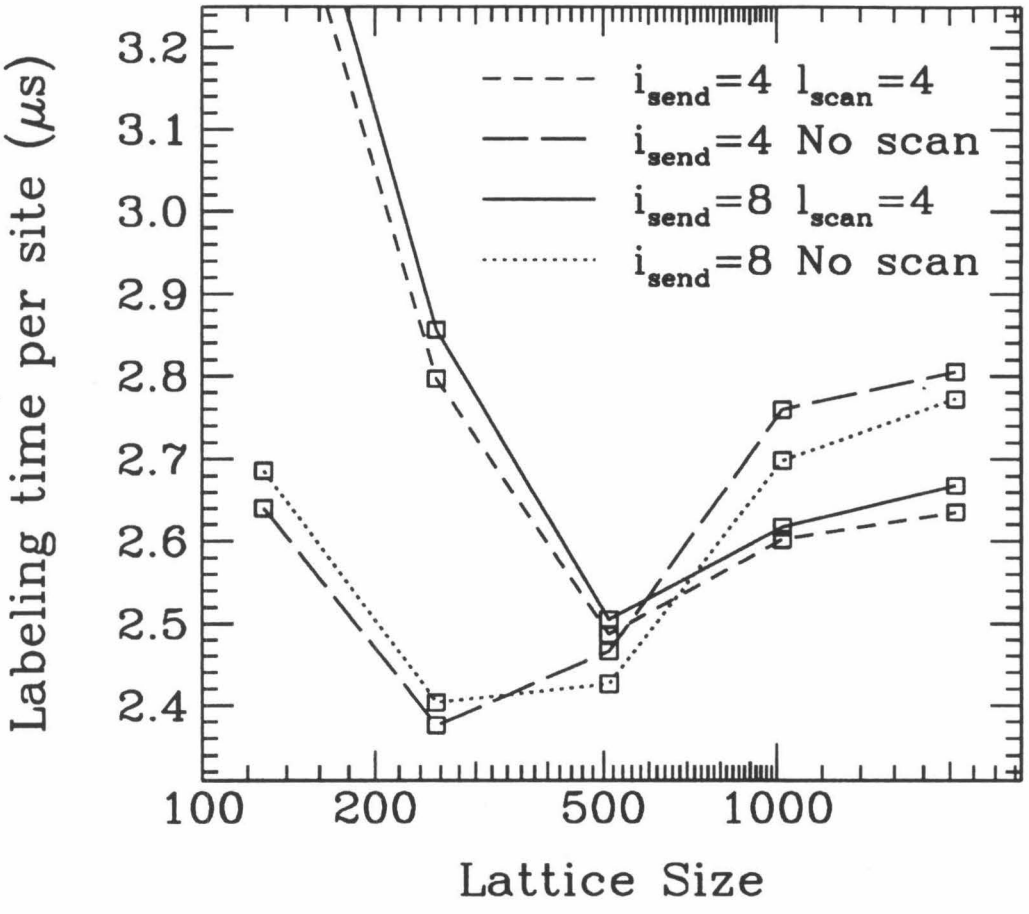

Figure 5.7: Average times per site as a function of lattice size for labeling Ising model clusters on the CM-2, for different values of the parameters for the *get/send* algorithm. i_{send} is the initial number of steps without a *get* or *send*. Here the interval between *get*s is $p_{get} = 1$, while $p_{send} = p_{scan} = 2$. l_{scan} denotes the number of steps after which *scan*s are not performed.

We also tried a very simple telescoping scheme in order to determine whether any benefit could be derived from reducing the size of the problem in such a way. For this we used only the top level and one lower level, consisting of the even-even coordinate sites, and attempted to solve this partial problem by iterating until the labels of this sub-lattice did not change. Neither this nor an attempt to do a fixed number of multi-scale iterations on the lower level managed to reduce the amount of time required for the full labeling procedure. We note that only a single reduction and expansion was tried. This poor result was not due to the overhead of communicating between the different lattices, which cost very little time, but must have been due to the lack of information at the lower level about enough of the important connections.

5.3 Get/send : An algorithm using general communications

We now present a different component labeling algorithm[80] that was inspired by the efficiency of the SIMD algorithm of Hillis and Steele[81] for finding the end of a linked list. We treat the labels as pointers in a dynamic tree-like structure, making our method similar to the sequential algorithms of Galler

and Fisher[82] and of Hoshen and Kopelman[83] and the parallel algorithm of Shiloach and Vishkin[84].

All the cluster labeling algorithms discussed to this point start with each site being given a unique label. It is convenient to set the original label to be the site number. We will again label a cluster by the smallest starting label of all the sites in that cluster. For a two-dimensional lattice we could assign the original cluster label C of the site (x, y) to be $(y * L) + x$, for example. During the cluster labeling procedure we can consider the current label as a pointer to the site where it originated (*i.e.*, $x_{orig} = C \bmod L$, $y_{orig} = C/L$). The site (x_{orig}, y_{orig}) started with the current label of the site (x, y), and as we shall see, at any later time in the labeling procedure it must have a label which is less than or equal to this value. Thus at any time each site can get the label of the origin (x_{orig}, y_{orig}) of its current label and use it as its new label. On the Connection Machine this is done using the general communication routine **get**. To ensure that the algorithm eventually gives the correct result, at each stage a local label propagation step must be performed as well.

By itself this method will correctly label any lattice but performs very badly, because in many places labels propagate only with the local step, on paths that can be very long. To overcome this problem we have added an

important supplement – an inverse step which propagates information large distances in the opposite direction and proceeds as follows. Each site saves its label, and then performs a local iteration. It then compares its current label with that old label and, if they are different, sends the current label to the originating site of its old label. For this step we use the Connection Machine routine **send_with_minimum**, for which any site that is sent more than one value keeps only the smallest. Each site then takes the minimum of the labels it is sent, if that is smaller than its current label.

Doing a *send* step before each *get* means that if the label of any site is changed, the new label is then propagated immediately (by the *get*) to all sites with the old label. Thus we can wholly relabel a large area, or sub-cluster, in one step as soon as it contacts another large sub-cluster with a smaller label. In Fig. 5.3 we show the average number of iterations required to label the Ising model clusters. The data fit perfectly to a logarithmic increase with the lattice size.

The costliest parts of this algorithm are the *get* and *send* steps, which require general communication routines which take about ten times longer than local grid communications on the CM-2. This is compensated of course by the value of the information which is passed over large distances. However a way of doing some of the work by a less costly method will reduce the total

running time. Since in the first few iterations the *get/send* step is able to do very little useful work, at a large cost, it is always faster to wait for i initial iterations before using it. Also, depending on the machine specifics and the efficiency with which *get* and *send* are implemented, it may be more efficient to do the *get/send* step only at every pth iteration of the algorithm, with just local label propagation for the other iterations. The parameters i and p can be tuned to optimize the algorithm for any particular application, system size, and parallel computer. We found that for our application on the CM-2, $p = 1$ or 2 and i between 4 and 8 generally gave the best results.

We also note that for large lattices it is not necessary to *get* the label of the originating site for every point since, in any sizable cluster, most neighboring points belong to the same cluster. We have experimented with having only a portion (we use a quarter) of the sites perform a *get*. The new labels received are then transferred to neighboring sites by the subsequent local iteration. Whether this modification proves to be the fastest option depends of course on the communications hardware and software of the particular machine. For the CM-2, the time for each *get* is roughly halved, although the number of iterations required is increased slightly, so there is a trade-off. For an earlier implementation of the *get* routine on the CM-2 this method was substantially faster; however, an improved *get* now means that it is about 10% faster to

do the *get* step at every site.

Another way of improving the performance of this algorithm is to combine it with the *scan* operation. Using *scan* in some of the initial iterations pushes the label more efficiently over moderate distances. On the CM-2, doing a *get* and a *send* is about as expensive as doing the 4 *scans* (one in each direction) of a *scan* step, however it provides a much better way to propagate labels by making large changes at ever increasing length scales, and of handling the large, irregular and labyrinthine clusters for which the *scan* algorithm fares poorly. Including a few initial *scan* steps makes the algorithm slightly more efficient at large VP ratios (*i.e.*, larger lattice sizes), but again this will be highly dependent on the specifics of the problem and the machine.

Of course we continue to use connection improvement for the local steps, and see a benefit for all the variations of the algorithm. Fig. 5.7 shows the average time per site to label lattices of different sizes. It is evident that between sizes of 128 and 512 the efficiency of large VP ratios reduces the average time, while for larger lattices this behavior subsides and it is dominated by the increase in the number of iterations needed to converge.

The average labeling time per site for the basic *get/send* algorithm ($i = 0$, $p = 0$, and no *scans*), as well as the optimized algorithm, is shown in Fig. 5.4. We can see that this method is substantially faster than the multi-

scale algorithm.

5.4 Discussion and conclusions

Our labeling algorithms are very general, and can be applied to any application where component labeling is necessary, such as percolation[85], image analysis[72], and for the various cluster Monte Carlo algorithms which have been proposed for many different spin models[68, 60, 70].

We have presented numerical evidence that the average number of iterations required by our algorithms to label percolation-like Swendsen-Wang clusters at the critical point of the Ising model, which are highly irregular in both shape and size, scales with the logarithm of the lattice size. Up to corrections caused by differing VP ratios on the Connection Machine, the times required for the labeling using these algorithms scale as $(\log_2 L)^2$ per lattice site.

There is only a subtle difference between the multi-scale and *scan* algorithms: both methods look at connections at distances 1, 2, 4, etc., but for multi-scale we also do a comparison at each distance and set the connection accordingly. This connection improvement is enough to give the multi-scale algorithm significantly better scaling behavior.

Our multi-scale algorithm is simpler than that of Brower *et al.*[76], and appears to scale better with increasing lattice size. For the lattice sizes of interest (of order 1024×1024), our optimized algorithm gives an average labeling time for the Ising problem of 6.5 microseconds per site on a 16K CM-2 running at 7 MHz, which is comparable with the time of 6.0 μs per site obtained by Brower *et al.* for the same size machine. However this time for the optimized *get/send* algorithm is substantially better, at 2.6 μs per site.

These kind of SIMD algorithms work quite well on massively parallel fine grained SIMD machines like the CM-2, as long as the objects to be labeled are fairly small, for example in image processing applications such as analyzing images on a radar screen. However fine grained SIMD parallelism does not usually work well for problems which are very irregular and require a lot of non-local communication. Unfortunately the clusters to be labeled in spin model and percolation applications are very large and irregularly shaped, and we would therefore expect that it would be very hard to get good performance for labeling algorithms on these problems using fine grained SIMD machines. This is reflected in our results, since cluster labeling for the Ising model can be done at a rate of about 5 μs per site on a single IBM RS/6000-550 workstation, compared to 2.6 μs per site with our best algorithm on a 16K

CM-2.

However we have previously obtained quite good efficiencies on coarse grained MIMD machines for parallel component labeling algorithms which use only local propagation of labels[73], and thus do not scale well for very large numbers of processors. Incorporating the above multi-scale and general communication (*get/send*) ideas into these MIMD algorithms promises to allow us to greatly improve their efficiency and scalability, and thus exploit the power of large MIMD parallel supercomputers such as the nCUBE, the Intel machines, and the CM-5.

Note added: After this work had been substantially completed, and preliminary results reported at a conference[80], we found that the *get/send* algorithm had been independently proposed by P. Rossi and G.P. Tecchiolli[86].

Chapter 6

The Renormalized Trajectory and the β Function.

As we have seen, Monte Carlo simulation of the O(3) model on large lattices has not been able to find asymptotic scaling for the standard action. Since the limits of current computational resources have been reached in the study on lattices large enough for the correlation to have its infinite volume value, in order to investigate the model further other techniques are needed. One of the most important techniques developed in recent years that can be profitably used to extend our range of investigation is the Monte Carlo renormalization group (MCRG).

A recent study of the model using an MCRG technique by Hasenfratz and

Niedermayer[19] measured the discrete beta function $\Delta\beta(\beta)$ for the standard action (SA) up to $\beta = 2.24$ (where $\xi \simeq 1090$) and for the tree-level improved action (TIA) up to $\beta = 1.92$ ($\xi \simeq 330$). The discrete beta function $\Delta\beta(\beta)$ is the difference between the coupling β and that coupling β' at which the correlation length ξ is half that at β, *i.e.*, $\Delta\beta(\beta) = \beta - \beta'$ where β' is chosen so that $\xi(\beta') = \frac{1}{2}\xi(\beta)$. Thus it tracks the behavior of the β function of the theory between $\beta - \Delta\beta(\beta)$ and β.

For the TIA they observed convergence of $\Delta\beta$ to the asymptotic scaling predictions (within their statistical errors) starting from $\xi \simeq 40$; from this they calculated the value of $m/\Lambda_{\overline{MS}} = 3.4(1)$. For the SA only the last point at $\beta = 2.26$ is compatible with asymptotic scaling; since the value obtained $m/\Lambda_{\overline{MS}} = 3.3(1)$ agreed with that for the TIA, they surmised that AS is probably achieved. However, the evidence is not convincing for the standard action because only one measurement of $\Delta\beta$ agrees with the prediction.

It was the same authors, together with Maggiore[18], that then calculated the exact value of $m/\Lambda_{\overline{MS}}$by relating the Bethe Ansatz solution of the model with perturbation theory. The value obtained, $m/\Lambda_{\overline{MS}} = 8/e \simeq 2.943$ is in obvious disagreement with the MCRG result. Since the numerical results are unconfirmed for $\beta > 2.1$ and there is a gap between the measurements and theory we have decided to undertake an investigation of the model using an

alternative MCRG technique. We will use this to check the previous calculation and extend the results further in the continuum limit of the theory.

In the subsequent sections we will, respectively, introduce the Monte Carlo Renormalization Group, review the methods for directly obtaining the couplings of the block Hamiltonians, present the equations of motion method, test the method and our implementation, present our results for the standard block transformation and for a probabilistic block transformation, and derive our conclusions.

6.1 The Monte Carlo renormalization group

The Renormalization Group (RG) is an important tool for the investigation of field theories. By studying the way in which thinning the degrees of freedom of a theory changes the Hamiltonian in the vicinity of a critical point, many properties of a theory can be extracted. Analytical renormalization group calculations have provided much insight into the behavior of many models, including, as we saw, the ultraviolet limit of the $O(N)$ nonlinear sigma model. However analytical calculations cannot be performed for all models and cannot be extended from the critical points to all regions of interest of a model. To fill this gap a powerful technique has been developed

that merges the strengths of the Renormalization Group with the flexibility of the Monte Carlo simulation methods[87, 88]. With this method, the Monte Carlo Renormalization Group (MCRG), a theory can be studied in regimes in which perturbation theory and other expansions cannot be applied and its behavior can either be independently explored or it can be tracked into a region where it is well understood.

The basic feature of a renormalization group study of a model is tracking the flow of the Hamiltonian under a change of scale. The degrees of freedom are merged or thinned by being integrated out. In the real space renormalization group this blocking is achieved by merging neighboring degrees of freedom, usually in a hypercube of length b, into a single field. The remaining degrees of freedom are governed by a new Hamiltonian. By repeated block transformations the Hamiltonian will move (in the space of possible Hamiltonians) in a manner which reveals the long distance behavior of the model. Thus for models with a second-order phase transition there is a fixed point that governs the critical properties of the theory.

Different block transformations can be used and will lead to different fixed points, all of which reside on the critical hyper-surface of the model. However all physical results, for example critical exponents, must be independent of the choice of transformation. We also note that an important

additional assumption that is made (which is required for the method to be practically useful) is that the fixed point Hamiltonian is dominated by short range interactions.

The MCRG methods most commonly used do not directly determine the couplings of the block Hamiltonian. Instead they involve either Swendsen's procedure to determine derivatives of the block couplings with respect to the previous level coupling constants[88, 89] or Wilson's technique[90] of matching correlation functions (or ratios of such functions) attempting to pair up points (Hamiltonians) which have correlation lengths that differ by a constant factor, usually two, in order to retrieve the discrete beta function of the model. The matching method requires that the original line of Hamiltonians chosen is close to the renormalized trajectory of the block transformation that is used. Only then can there be agreement between the correlation functions measured at corresponding blocking levels.

Most MCRG calculations have been done using a variant of this technique: in the first type an action must be chosen so that it is close to the renormalized trajectory of the chosen block transformation, as proposed by Wilson[90] and utilized and extended by Shenker and Tobochnik[3], Hirsh and Shenker[91] and Hasenfratz and Margaritis[11]. In the second type a line of actions is chosen and the block transformation is adjusted such that the

renormalized trajectory is close to this action, as proposed by Hasenfratz *et al.* [30] and Swendsen[92]. For a review of many of these methods see, e.g., R. Gupta[93].

6.2 Determining the Block Hamiltonian

The matching techniques work well in the vicinity of the fixed point and the renormalized trajectory. However, as they depend on the chosen action being close to the renormalized trajectory of the transformation, they cannot be used for arbitrary values of the couplings. Thus an assumption has to made at the outset that the line of actions is close to the renormalized trajectory, and this must be validated afterwards.

What is needed in order to get around these restrictions is a method for directly extracting the couplings of the renormalized Hamiltonian. Then the effect of the block spin transformation can be studied in detail and the renormalized trajectory can be found in a straightforward way.

More powerful techniques do exist that directly determine the coupling constants of the block Hamiltonians. At least three practical techniques have been proposed that attempt to do this for an arbitrary renormalization group transformation (RGT). The first is based on the Schwinger-Dyson

equations, another relies on a variant of the microcanonical simulation, while a third uses a trial block Hamiltonian and an additional way of measuring correlation functions (that uses it) to iterate to the correct value. These methods, somewhat surprisingly, have remained largely unused and to our knowledge there have been no comparative tests of their relative efficacy, especially their different sources of systematic errors, the statistical errors of the resulting estimates, their robustness and the costs involved in using them. Preliminary results, though, are encouraging for all these methods.

In particular, the first method determines the coupling constants of block Hamiltonians using the Schwinger-Dyson equations[94]. These can be expressed as equations of motion for different correlation functions, e.g., for two-spin dot products like the nearest or next nearest neighbor spins. They equate expectation values of two-point and higher-order correlation functions and of products of spins and derivatives of the Hamiltonian. By truncating the range of the interactions and substituting the coefficients of the unknown Hamiltonian, a simple linear system of equations is retrieved. This is solved to obtain the values of the relevant coupling constants.

The next method is based on an extension to the microcanonical simulation method of Creutz[95] known as Creutz's demon. Creutz's demon is a Monte Carlo method that relies on an extra degree of freedom, called a

demon. The demon travels around the lattice updating the spins at each site such that the sum of the energy of the lattice and the demon remains constant and the energy of the demon always remains positive. The value of the coupling is not an input parameter of the simulation, but is derived as a result of the run; the estimate of the energy provides the data from which the value of the coupling constant can be extracted. By introducing one demon for each of the correlation functions of interest (and using an auxiliary simulation for each blocking level of interest) M. Creutz, A. Gocksch, M. Ogilvie, and M. Okawa[96] obtained an estimate for the corresponding coupling constants.

Swendsen's method was actually the first one proposed[97]. It uses a trial block Hamiltonian and an additional estimator for correlation functions that depends on that trial Hamiltonian. In an iterative process, the two estimates for each of several observables are compared for the same simulation, and the differences provide a correction for the Hamiltonian.

6.3 The equations of motion method

We use the method developed by M. Falcioni, G. Martinelli, M.L. Paciello, G. Parisi, and B. Taglienti[94]. It is based on Schwinger-Dyson equations, so we have called it the "equations of motion method." It computes the

couplings of the block renormalized Hamiltonians and can thus be used to extract the trajectory of an arbitrary renormalization group transformation, starting from any point (initial Hamiltonian).

We have obtained a more general and compact formulation of the equations of motion of any observable by using the Langevin equation. For an observable A, we consider its derivatives

$$A_{,i}^{\ a} = \frac{\partial A}{\partial \sigma_i^a} \tag{6.1}$$

(the lower index i is the site index and the upper indices are the spin component labels) and the generalized force derived from the Hamiltonian

$$F_i^a = -\frac{\partial H}{\partial \sigma_i^a}. \tag{6.2}$$

The derivatives in the previous equations are taken without considering the constraint on the spins $\sum_a (\sigma_i^a)^2 = 1$. In order to formulate the equation of motion compactly we introduce a projection operator

$$P_i^{ab} = \delta^{ab} - \sigma_i^a \sigma_i^b \tag{6.3}$$

The lower site index i is not summed over in the above or subsequent equa-

tions; the upper indices (the spin component labels), however, are summed over when repeated. The equation of motion for the quantity A can be obtained by imposing the condition that its average is invariant under $O(N)$ rotations of the fields and is derived in appendix B. It is

$$A^{\ a}_{,i} P^{ab}_i F^b_i + P^{cd}_i A^{\ cd}_{,ii} = (N-1)\sigma^a_i A^{\ a}_{,i} \tag{6.4}$$

where the equality is understood to refer to averages of the correlation functions.

We can consider the equations of motion for observables which span a set of short range correlation functions, for example the nearest neighbor correlation function and the diagonal one $(1,1)$. If we use a truncated Hamiltonian H with short range couplings we obtain a system of linear equations in the couplings β_a of H. The simplest useful observable is the two-spin correlation function $A = \vec{\sigma}_0 \cdot \vec{\sigma}_j$, with $j \neq 0$ any lattice point. For this the equation of motion becomes

$$\sigma^a_j(\delta^{ab} - \sigma^a_0\sigma^b_0)F^b_0 = (N-1)\sigma^a_0\sigma^a_j. \tag{6.5}$$

If we take a Hamiltonian with two terms, the nearest neighbor and first

diagonal term

$$H = -\beta_{NN} \sum_{\substack{i \\ \hat{\mu}=\pm\hat{x},\hat{y}}} \vec{\sigma}_i \cdot \vec{\sigma}_{i+\hat{\mu}} - \beta_d \sum_{\substack{i \\ \hat{\nu}=\pm\hat{y}}} \vec{\sigma}_i \cdot \vec{\sigma}_{i+\hat{x}+\hat{\nu}}, \tag{6.6}$$

where the index i runs over all sites, we can derive the force

$$F_0^a = \beta_{NN} \sum_{\substack{i \\ \hat{\mu}=\pm\hat{x},\pm\hat{y}}} \sigma^a_{i+\hat{\mu}} + \beta_d \sum_{\substack{j \\ \hat{\mu}=\pm\hat{x}\pm\hat{y}}} \sigma^a_{j+\hat{\mu}} \tag{6.7}$$

and substitute it into equation (6.5). We obtain equations (one for each point j) for the nearest neighbor and diagonal couplings

$$\begin{aligned} \beta_{NN} \sum_{\vec{u}\in V_{NN}} [(\vec{\sigma_0} \cdot \vec{\sigma_j}) - (\vec{\sigma_0} \cdot \vec{\sigma_j})(\vec{\sigma_j} \cdot \vec{\sigma}_{\vec{u}})] + \\ \beta_d \sum_{\vec{u}\in V_d} [(\vec{\sigma_0} \cdot \vec{\sigma_j}) - (\vec{\sigma_0} \cdot \vec{\sigma_j})(\vec{\sigma_j} \cdot \vec{\sigma}_{\vec{u}})] \\ = N(\vec{\sigma_0} \cdot \vec{\sigma_j}) - 1 \end{aligned}$$

where the sets of vectors are $V_{NN} = \{\hat{x}, \hat{y}, -\hat{x}, -\hat{y}\}$ and $V_d = \{\hat{x}+\hat{y}, \hat{x}-\hat{y}, -\hat{x}+\hat{y}, -\hat{x}-\hat{y}\}$. Thus using the above equation for at least two points j we have a system of linear equations in the couplings β_{NN} and β_d, which we can solve to obtain estimates of their values.

In this case, and in general, the coefficients of the system of linear equations and the inhomogeneous term (or source vector, *i.e.*, b of $A \cdot x = b$) are

sums of correlation functions, which can be evaluated efficiently using Monte Carlo simulation.

We also note that equation (6.5) is the same as that is obtained by the approach of Falcioni *et al.* [94]. Our approach however simplifies its expression and is much better suited for calculating the equations for more complex observables and Hamiltonians.

For a more realistic case than our simple example more couplings must be included. As many as possible of the relevant interactions and those that are generated with large coefficients near the chosen starting points should be used, in order to obtain reliable and meaningful results. In this model this means that we must use the two point interaction for the nearest neighbor (NN), diagonal (d), and next nearest neighbor (NNN) [and where necessary and possible the third-nearest neighbor $(3,0)$] points and the shortest range four point interactions. These should include the four-spin couplings of the type $(\vec{s}_i \cdot \vec{s}_j)^2$ that couple to the square of the correlation function used in each of the above two-spin interactions, and some that involve spins at three and four neighboring sites, particularly $(\vec{s}_0 \cdot \vec{s}_{\hat{x}})(\vec{s}_0 \cdot \vec{s}_{\hat{y}})$ and $(\vec{s}_0 \cdot \vec{s}_{\hat{x}})(\vec{s}_{\hat{y}} \cdot \vec{s}_{\hat{y}+\hat{x}})$. Unfortunately the last two examples introduced too much complication for the present study, as they would have required measuring several large new classes of correlation functions beyond those we currently measure.

We obviously require at least as many equations as there are coupling constants to find. We also note that simple observables, which are symmetric in their endpoints, generate only one equation. Other unsymmetric observables (where the endpoints are not all interchangeable, for example $\vec{\sigma}_0 \cdot \vec{\sigma}_{\hat{x}} \quad \vec{\sigma}_0 \cdot \vec{\sigma}_{\hat{y}}$) generate two or more. Thus in general we will need at least as many observables as equations.

In the Hamiltonian we chose to include all symmetric bilinear terms of distance less than r and the square of all such terms. We chose $r = 2$ and in some cases $r = 3$. Thus the Hamiltonian we use is

$$H = -\sum_{\alpha} \beta_\alpha \sum_{\substack{i \\ \vec{u} \in T_\alpha}} \vec{\sigma}_i \cdot \vec{\sigma}_{i+\vec{u}} - \sum_{\alpha} \gamma_\alpha \sum_{\substack{i \\ \vec{u} \in T_\alpha}} (\vec{\sigma}_i \cdot \vec{\sigma}_{i+\vec{u}})^2 \tag{6.8}$$

where α is an index running over the vectors that characterize the couplings - $(1,0),(1,1),(2,0),(2,1),(2,2)$ for $r = 2$ - and T_α is the set of vector that are the reflections and rotations of the α modulo the rotation by π, *i.e.*, for a vector $\alpha = (i,j)$ where $i \neq j$, T_α is the set $\{(i,\pm j),(\pm j,i)\}$.

We also chose to use as observables A in 6.4 the set of bilinear correlation functions $\vec{\sigma}_0 \cdot \vec{\sigma}_i$,and of their squares $(\vec{\sigma}_0 \cdot \vec{\sigma}_i)^2$ with i visiting all possible sites whose coordinates have an absolute value not greater than l. More precisely $i \in \{(s,t)/0 \leq s \leq t \leq l \text{ and } a > 0\}$. We made the choice $l = 2$ in most cases

because of the large computational cost involved in choosing higher values, but also ran tests with $l = 3$.

Having chosen the trial Hamiltonian and the set of observables we measure, for each blocking level of interest, the set of correlation functions required to determine all the coefficients and the inhomogeneous terms of the system of linear equations.

Let us also briefly describe the method of Creutz *et al.* [96] for comparison purposes. To use it first one measures the correlation functions corresponding to the coupling constants desired for a set of block lattices (e.g., for β_{NN} we would measure $< \vec{s}_i \cdot (\vec{s}_{i+\hat{x}} + \vec{s}_{i+\hat{y}} >$). Then a lattice with initial values of these correlation functions H_i equal to the mean $E_{i,0}$ of the measurements is prepared. An extra degree of freedom, called a demon, is then introduced for each correlation function i and is allowed to take values E_D^i in a restricted range. An auxiliary simulation is then run with the partition function

$$Z_{\mu C} = \sum_{-1<E_D^i<1} \sum_{S_n} \prod_i \delta\left(H_i(S) + E_D^i - E_{i,0}\right) \tag{6.9}$$

where $H_i(S)$ is the value of the i correlation function that is coupled to the coupling constant β_i. From the average value $< E_D^i >$ of the i demon's "energy" one can extract the value of the coupling constant β_i by using the

following equality

$$< E_D^i >= \frac{1}{\beta_i} - \frac{1}{\tanh \beta_i}. \tag{6.10}$$

This is valid in the large volume limit, so in some cases finite volume corrections are needed. We note that this method is not ergodic, so checks must be made whether the results are reliable. However in all cases to date this has not presented a difficulty.

This technique is unfortunately less well suited to parallel computers. To preserve the method it seems necessary to simulate it exactly as if it is run sequentially. Satisfying the constraints necessary to do this in a parallel environment would make the method fairly inefficient.

6.4 Testing the equations of motion method

We performed several tests in order to measure the accuracy of the method and to test the correctness of our implementation. But before we show results of such a comparison we will mention a few details of our simulation.

We used a cluster Monte Carlo update algorithm; in particular we used the Wolff method[29] to map the O(3) spin model onto a ferromagnetic Ising spin model. However we did not use his single cluster dynamics, in which only one cluster is built at a time starting from a random point on the lattice

and that cluster is always flipped. We chose instead to use the Swendsen-Wang dynamics[28] which identifies all clusters and flips each cluster with probability 1/2. We did this because we believe it makes better use of the parallel computer we used, the Connection Machine-2 from the Thinking Machines Corporation.

To check how fast the algorithm produces new configurations we measured and saved the total spin of the lattice and the energy after every sweep. We observed that the integrated autocorrelation time of the magnetizations is $\tau_{int}^{M} \simeq 2.5$, *i.e.*, in all cases it remains small. Measurement of the appropriate correlation functions was made every 20 sweeps, so that we measure these quantities on configuration that are correlated little. In this way the update consumes about 20 per cent of the CPU time. Correlation functions are summed over a certain number of iterations, 2500 for $L = 256$, 1500 for $L = 512$ and 500 for $L = 1024$ and saved. They are then read in by a separate analysis program which computes and solves the system of linear equations for each set saved. Errors are obtained by averaging the measurements and also by using the jackknife method in the following way: the possible sets of all but one correlation function measurements are constructed, the linear system of equations are solved for each set and the jackknife error equation is used to recover the error estimates[56]. The two methods give similar errors,

β_{orig}	β_{NN}	β_d	β_{NNN}	$\beta_{(2,1)}$	$\beta_{(2,2)}$
2.0	2.0010(15)	0.0018(23)	-0.0016(16)	0.0022(6)	0.0002(21)

$\beta_{(3,0)}$	$\beta_{(3,1)}$	$\beta_{(3,2)}$	$\beta_{(3,3)}$
-0.0034(12)	-0.0016(8)	0.0011(15)	-0.0013(18)

Table 6.1: Direct test of the method using no blocking and running at $\beta = 2.0$. Couplings extracted using the Schwinger-Dyson equations with only bilinear couplings in the Hamiltonian are shown. We included all interactions in an orthogonal triangle with small sides of length $r = 3$ and one point at the origin (0,0).

and we use the latter ones.

We used our method on a sample of configurations generated by Monte Carlo simulation without doing any blocking. Obviously the original couplings should be recovered in this case, allowing us to check the accuracy and robustness of our procedure. Table 6.1 shows the results of such a test using only 750 measurements of all appropriate correlations functions. The agreement with the input values of the simulations and small estimated errors show that we can faithfully extract the value of the couplings for these cases and that statistical errors are under control in our estimates.

The couplings obtained are obviously not statistically independent. They have a complicated covariance matrix. We cannot show all the elements of this nine by nine (and in other cases ten by ten or larger) symmetric matrix for each case because it would take too much space to do so. Instead we will quote the error derived from just the diagonal parts of the covariance

matrix for each entry in our tables. We do use the full matrix, however, in appropriate parts of our subsequent analysis. We will note that the values of the linear correlation coefficients derived from the covariances take on similar values for the same entry on different lattices and levels. Different entries exhibit almost the full range of possible values, from close to -1 to about 0.8.

As a further check, we measure the couplings at each blocking level. Then we simulate a Hamiltonian with couplings equal to our estimates of the block Hamiltonian at the first blocking level on a lattice half the size. We compare the measurements (at one level less) in this simulation with the estimates on the original lattice.

Table 6.2 shows the results of the above tests starting at $\beta = 2.0$. We see that matching between corresponding levels is reasonable but not very good - other terms in the Hamiltonian (in particular the quadratic terms with the γ couplings, the third neighbor bilinear coupling β_{NNN} and possible other quadratic spin couplings). A simulation with these γ couplings included in the second Hamiltonian and in the analysis of both runs is needed to determine whether agreement would be much improved by their inclusion. This has not yet been possible. We also see that including the γ couplings in just the second simulation changes those results substantially.

L	lv	β_{NN}	β_d	β_{NNN}	$\beta_{(2,1)}$	$\beta_{(2,2)}$
1024	0	2.0				
1024	1	3.2342(8)	-0.0105(8)	-0.3985(10)	-0.0436(5)	0.0184(9)
1024	2	3.9654(16)	-0.1963(25)	-0.6758(23)	-0.0337(17)	0.0487(27)
1024	3	3.9563(36)	-0.2806(62)	-0.7266(50)	-0.0168(34)	0.0510(54)
512	0	3.234	-0.011	-0.398	-0.044	0.018
512	1	4.0185(10)	-0.1554(22)	-0.7899(18)	-0.0379(14)	0.0643(17)
512	2	3.7190(23)	-0.2512(31)	-0.7284(19)	-0.0156(22)	0.0562(28)
512	3	2.6176(46)	-0.1578(64)	-0.4702(56)	-0.0165(40)	0.0257(47)
512	4	1.9779(84)	-0.0739(84)	-0.3504(83)	-0.0149(72)	0.0124(74)
512	1′	4.908(6)	-0.211(6)	-0.986(2)	-0.044(2)	0.080(4)
512	2′	4.734(9)	-0.367(10)	-0.940(4)	-0.015(4)	0.065(6)
512	3′	3.042(11)	-0.207(15)	-0.550(8)	-0.014(7)	0.026(7)
512	4′	2.182(14)	-0.077(17)	-0.380(10)	-0.015(9)	0.014(9)
L	lv	γ_{NN}	γ_d	γ_{NNN}	$\gamma_{(2,1)}$	$\gamma_{(2,2)}$
512	1′	-0.629(4)	0.027(4)	0.171(2)	0.015(2)	-0.015(4)
512	2′	-0.740(7)	0.067(7)	0.198(3)	0.013(3)	-0.007(4)
512	3′	-0.375(9)	0.029(12)	0.113(8)	0.010(6)	-0.002(9)
512	4′	-0.233(12)	-0.007(15)	0.074(8)	0.002(7)	0.010(10)

Table 6.2: Coupling estimates at different levels for two simulations: one at *beta* = 2.0 on a 1024 lattice and the other on a 512 lattice at the coupling estimates of the first level of the previous one. Statistics are 13x500 for *beta* = 2.00 at $L = 1024$ and 36x1500 for $L = 512$. For the latter lattice we also show for comparison the couplings obtained when square bilinear terms are included in the analysis (levels 1′, 2′, ...)

By also measuring the correlation length we can establish whether the most significant part of the physics is captured when we cut off the couplings. Finding $\xi = 67 \pm 4$ for the simulation starting from the estimated couplings of level one of the run at $\beta = 2.0$ we see that it is much less than half the correlation length for $\beta = 2.0$. So using only these bilinear couplings neglects important effects. For our extensive runs we added the quadratic spin couplings $\gamma\alpha$ as well; a full check for this case has not yet been possible, as we mentioned above.

Since truncation errors seem to be significant we have tried to estimate their effect. We have done so by seeing how the neglect of terms similar to those used, but covering longer distances, generate systematic error. In order to do this we measured the correlation functions needed to determine the Hamiltonian truncated to $r = 3$ for runs with the standard action at $\beta = 1.4$. We then fitted to Hamiltonians with $r = 2$ and $r = 3$. The results can be seen in table 6.3. They show that couplings are stable, and others have a difference in a number of couplings of about 0.02, which is larger than the statistical errors. This gives us an estimate of the systematic errors because of truncation for coupling estimates using this renormalization group transformation.

Truncation errors are unavoidable in most methods. Even methods that

Hamiltonian with radius = 2 in the analysis, correlation functions with R=l+r=4

lv	r	l	β_{nn}	β_d	β_{nnn}	$\beta_{2,1}$	$\beta_{2,2}$	γ_{nn}	γ_d	γ_{nnn}	$\gamma_{2,1}$	$\gamma_{2,2}$
1	2	2	1.8384(6)	0.0374(6)	-0.2282(4)	-0.0340(3)	0.0075(3)	-0.0334(6)	-0.0057(8)	0.0352(5)	0.0098(2)	0.0012(6)
2	2	2	1.7428(12)	0.0021(11)	-0.2685(6)	-0.0321(4)	0.0115(6)	-0.0804(9)	-0.0058(12)	0.0393(8)	0.0078(5)	-0.0008(7)
3	2	2	1.2031(10)	-0.0003(8)	-0.1796(9)	-0.0212(6)	0.0066(6)	-0.0720(9)	-0.0114(11)	0.0148(8)	0.0027(8)	-0.0016(11)
4	2	2	0.6465(10)	-0.0009(7)	-0.0848(11)	-0.0090(9)	0.0040(8)	-0.0328(12)	-0.0048(12)	0.0016(13)	-0.0024(13)	-0.0002(15)

Hamiltonian with radius r=2 in the analysis, correlation functions with R=6

lv	r	l	β_{nn}	β_d	β_{nnn}	$\beta_{2,1}$	$\beta_{2,2}$	γ_{nn}	γ_d	γ_{nnn}	$\gamma_{2,1}$	$\gamma_{2,2}$
1	2	4	1.8343(16)	0.0261(15)	-0.2237(12)	-0.0243(9)	0.0035(9)	-0.0328(14)	-0.0036(13)	0.0347(11)	0.0068(7)	0.0027(12)
2	2	4	1.7393(28)	-0.0111(24)	-0.2603(23)	-0.0208(22)	0.0075(17)	-0.0808(29)	-0.0081(36)	0.0345(23)	0.0060(19)	0.0007(18)
3	2	4	1.2040(35)	-0.0060(26)	-0.1787(26)	-0.0159(24)	0.0031(21)	-0.0736(32)	-0.0067(31)	0.0190(28)	0.0060(35)	0.0026(33)
4	2	4	0.6489(38)	-0.0076(24)	-0.0807(33)	-0.0075(20)	0.0017(30)	-0.0299(48)	0.0008(49)	0.0050(32)	0.0005(35)	0.0075(37)

Hamiltonian with Radius = 3 in the analysis,

lv	r	l	β_{nn}	β_d	β_{nnn}	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{3,0}$	$\beta_{3,1}$	$\beta_{3,2}$	$\beta_{3,3}$
1	3	3	1.8384(15)	0.0412(14)	-0.2417(12)	-0.0404(7)	0.0009(12)	0.0262(12)	0.0066(10)	-0.0003(6)	0.0002(10)
2	3	3	1.7432(22)	0.0108(21)	-0.2890(18)	-0.0430(17)	0.0096(21)	0.0418(19)	0.0090(16)	-0.0018(14)	-0.0002(16)
3	3	3	1.2066(35)	0.0037(25)	-0.1965(21)	-0.0277(23)	0.0061(24)	0.0311(26)	0.0056(12)	-0.0002(17)	-0.0033(20)
4	3	3	0.6499(40)	-0.0068(25)	-0.0860(37)	-0.0081(21)	0.0022(26)	0.0147(36)	-0.0014(21)	-0.0010(22)	0.0027(24)

lv	r	l	γ_{nn}	γ_d	γ_{nnn}	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{3,0}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{3,3}$
1	3	3	-0.0322(14)	-0.0057(14)	0.0362(10)	0.0097(7)	0.0027(9)	-0.0005(10)	-0.0016(10)	0.0005(7)	0.0004(10)
2	3	3	-0.0763(25)	-0.0103(32)	0.0365(22)	0.0092(15)	0.0003(19)	-0.0045(15)	-0.0023(14)	0.0007(18)	-0.0019(15)
3	3	3	-0.0700(28)	-0.0073(33)	0.0186(26)	0.0069(34)	0.0018(32)	0.0013(31)	0.0043(26)	0.0010(21)	0.0028(23)
4	3	3	-0.0290(47)	0.0009(50)	0.0052(32)	0.0013(36)	0.0073(36)	0.0079(45)	0.0078(47)	0.0035(46)	0.0076(49)

Table 6.3: Comparison of couplings of block Hamiltonians at blocking levels 1-4 (lv) for different choices of truncations and sets of observables used to obtain them. The runs are at $\beta = 1.4$ on a 256^2 lattice. The Hamiltonians include all quadratic couplings β_i (that correspond to two spin correlation functions) and the quartic couplings γ_i as defined in the text. Only those which have and endpoint at (0,0) and at another point inside or on a triangle of side r are used [e.g., if $r = 1$, the points would be (1,0) and (1,1) while for $r = 2$ it also includes (2,0), (2,1) and (2,2)] . r here takes the values $r = 2$ and 3. All similar observables A within a similar triangle of side l are used for to obtain Schwinger-Dyson equations the linear system of which was solved to obtain the couplings shown. This neccesitated measuring all two, most four and many six point functions with endpoints in a box extending $R = l + r$ each side from a lattice point.

do not involve explicit truncation of the Hamiltonian, like the microcanonical method, including different couplings produces different results. These shifts have, for example, been seen in a simulation of the SU(2) gauge theory in four dimensions using the microcanonical method[98]. It is hard to compare the magnitude of these effects for different methods without doing an extensive comparative study, which is not possible in the current work. However the magnitude of these systematic errors seems to be no worse for the method we use than for other methods.

We should also point out that it is important to find the smallest set of correlation functions, which include all those needed and is also regular and thus simple to measure. The reason for this is the correlation functions of order four and six that are needed are those with endpoints at $(0,0)$ and two other points with coordinates of absolute value less than l, *i.e.*, $r = (x,y) \ : \ |x|,|y| < l$. Their number is proportional to l^4, which grows very fast with increasing l. By appropriate choice we have limited the number of correlation functions (other than the two-point functions) measured from $4(4r+1)^2[(2r+1)(r+1)-1]$ to $4(3r+1)^2[(r+1)(r+2)/2-1]$, which represents a significant decrease in the computation cost, *i.e.*, even for $r = 2$ it goes from 4536 to 980.

Lastly we note that the smallest lattices we take measurements on are

16^2. On smaller lattices the correlation functions we measure will obviously be dominated by finite size effects. Also, since the statistical errors grow much larger as the depth increases, we judged that measuring on smaller lattices would not be worthwhile.

6.5 Results for the standard RG transformation.

It is natural as a first case to study the simplest RG transformation. In this the block spin is chosen parallel to the sum $\vec{S}$ of the spins in a 2×2 block:

$$\vec{b} = \frac{\vec{S}}{|\vec{S}|}$$

We do this for intermediate to large couplings of the Standard action, $1.6 \leq \beta \leq 2.6$.

Our results for intermediate couplings, $1.6 \leq \beta \leq 2.0$, which are the most precise, can be seen in figure 6.1. Most couplings at first increase in magnitude after the first (and sometimes) block transformation. Then all converge to a single curve, the renormalized trajectory and proceed to follow it, decreasing in magnitude. This behavior and the renormalized trajectory

in this range of couplings and for this transformation were first observed by Okawa[99], using the microcanonical method of Creutz *et al.* [96]. He started from some standard action points, but found best convergence from two points with the block spin renormalization group improved action of Shenker and Tobochnik[3]. Comparing our results, in figure 6.1 we see that the trajectories are close but not overlapping. The broad similarity lends some credence to both methods, which are very different. The differences in the lines are probably due either to the truncation errors in both methods, or partly due to the finite effects in Okawa's measurements[99] which used 128^2 lattices.

Our confidence that we have found the renormalized trajectory is bolstered by the fact that simulating the action of Shenker and Tobochnik [3], our blocking steps also converge to the same curve. This can also be seen in figure 6.1.

Figure 6.1 shows that after three iterations of the block transformation we reach the renormalized trajectory. With further blocking steps the coupling constants remain on a single trajectory, evidence that we have found the renormalized trajectory.

What happens for larger values of the initial couplings β for the standard action can be seen in figure 6.2. Although there are larger errors in the

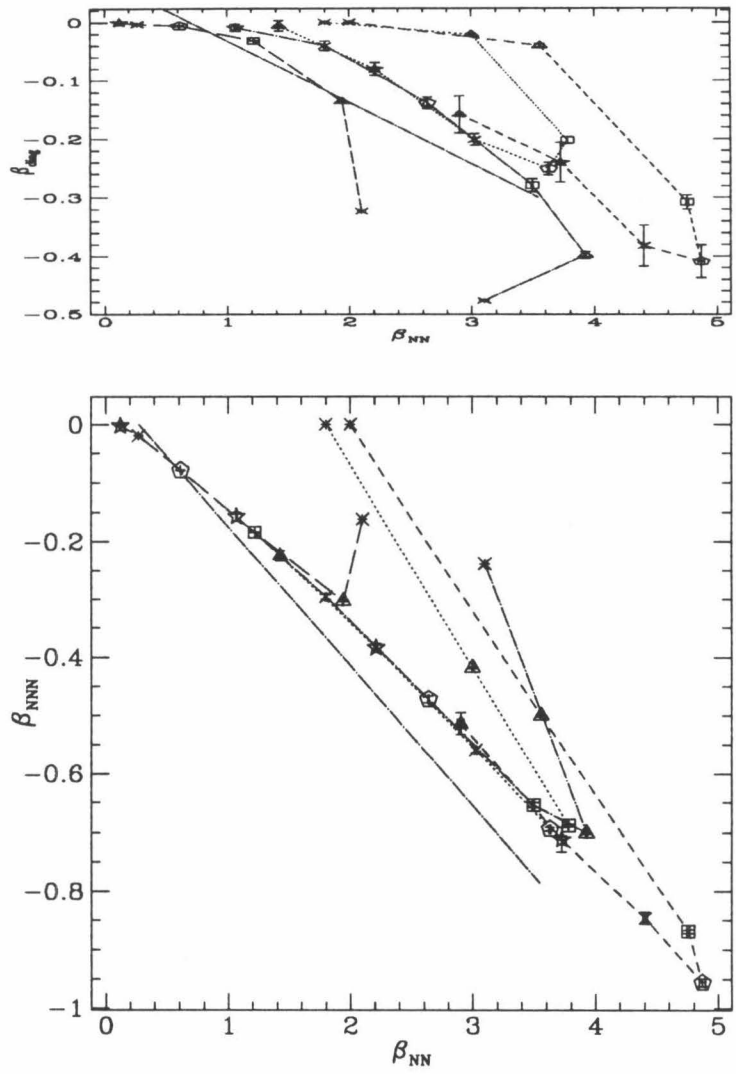

Figure 6.1: The coupling constants for the different blocking levels starting from the standard action and that of Shenker and Tobochnik using the standard RGT. Convergence to a single line, the renormalized trajectory, is clear. The line is the renormalized trajectory obtained by Okawa. Shown are the next nearest neighbor coupling and the diagonal coupling versus the nearest neighbor coupling. Different blocking levels are distinguished by their symbols.

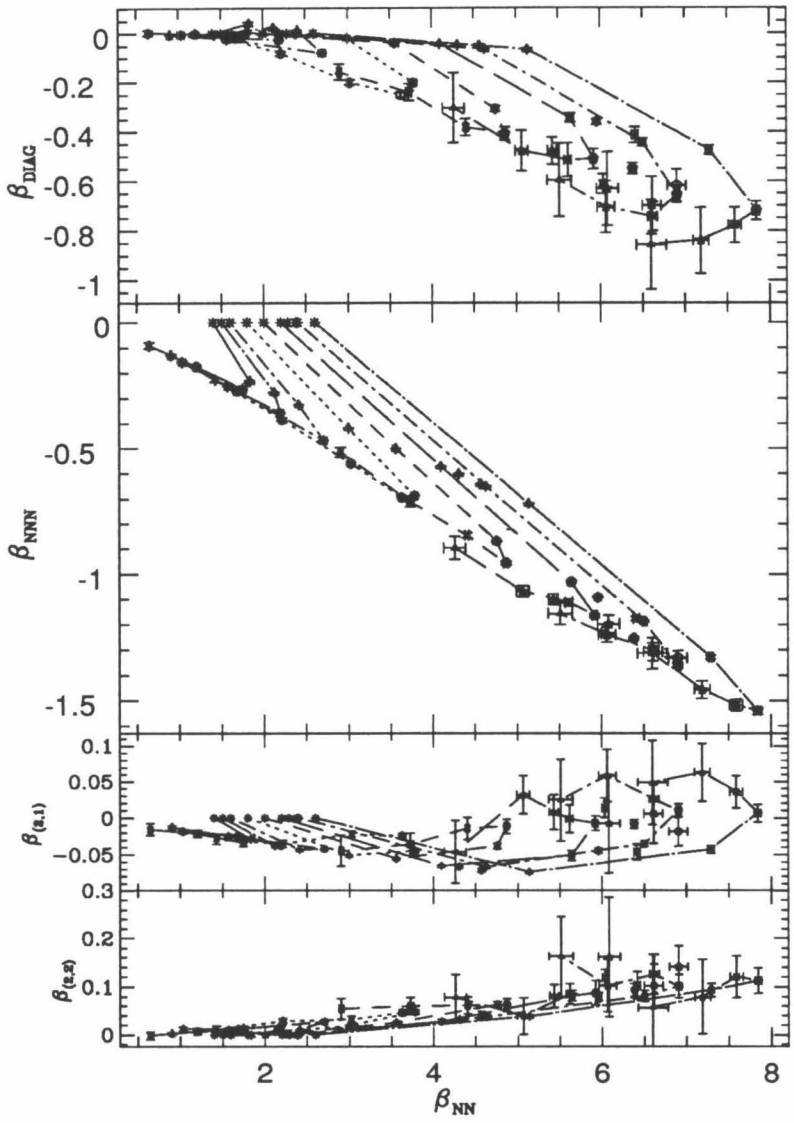

Figure 6.2: Flow of bilinear couplings under standard block transformation starting from several points of the standard, nearest neighbor action.

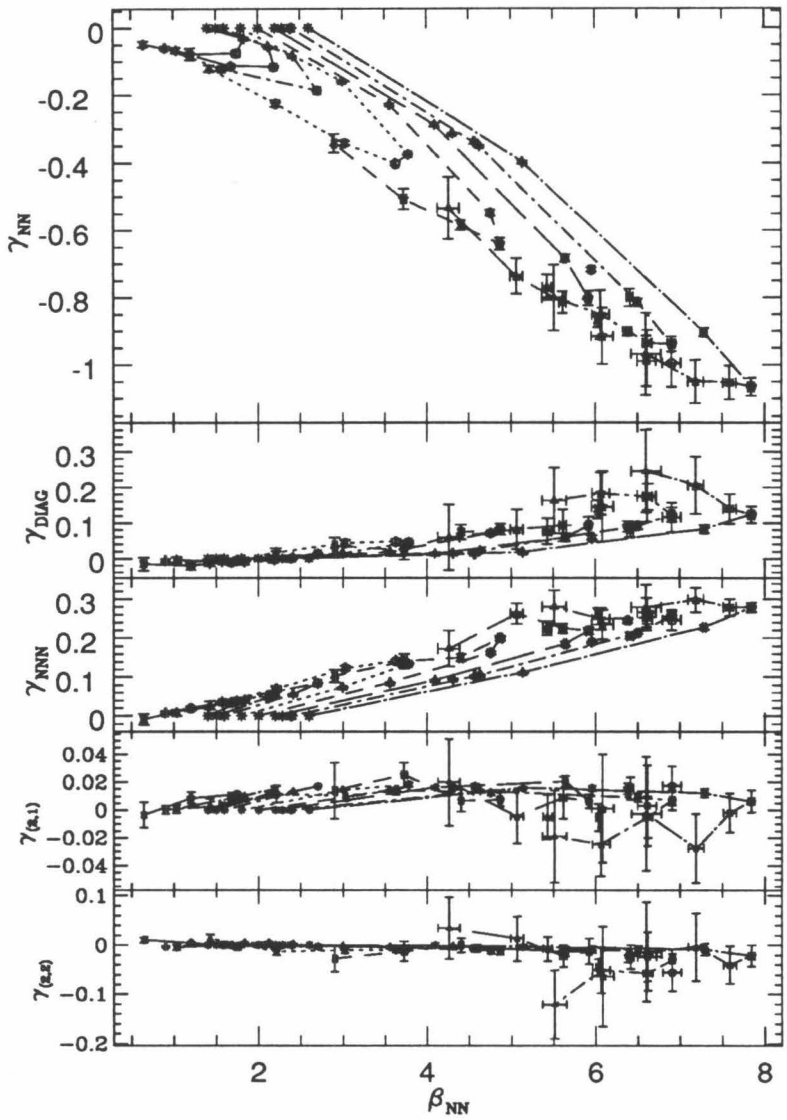

Figure 6.3: Flow of quartic couplings γ under standard block transformation. Same as previous figure, for different couplings.

estimates for $\beta \geq 2.2$, the convergence to a single curve that is the extension of the one seen at lower values of β is clear - for those coupling constants for which the value is much larger than the statistical error.

6.6 The improved RG transformation.

Another way to investigate the model is to vary the block transformation in order to obtain one that has a renormalized trajectory near the standard action. Hasenfratz *et al.* [30] suggested an improved renormalization group transformation. In this block transformation the block spin $\vec{b}$ is chosen from a probability distribution weighted by the dot product with the sum $\vec{S}$ of the spins in the block:

$$P(\vec{b}) = A(C) \quad \exp\left(-C\vec{b}\cdot\vec{S}\right) \tag{6.11}$$

where the parameter C is arbitrary; A(C) is chosen to provide the correct normalization. However C must be chosen so that the transformation has a fixed point. This dictates that $C \to \infty$ as $\beta \to \infty$, which makes the choice $C = c\beta$ a natural one. Hasenfratz *et al.* [30] used perturbation theory to obtain the optimal value for c for the $O(N)_{N\to\infty}$ model, so that matching could be possible at low blocking levels. In particular they observed that the continuum value of the beta function $\Delta\beta = \frac{\ln 2}{2\pi}$ was found by matching the

third and second blocking levels (3/2 matching) using the nearest neighbor correlation function and then also for 4/3 matching. (This comparison was made in the range $0.6 \leq \beta \leq 2.0$ corresponding to correlation length of $10 \leq \xi \leq 30000$.)

Hasenfratz and Margaritis[11] also used this method for the O(3) model and adopted the tree-level optimal value $c_{opt} = 2.3$. They were able to observe the change in the beta function from the crossover region $\beta < 1.6$ where it is larger than the two-loop result

$$\Delta\beta_{2loop}(\beta) = \frac{\ln 2}{2\pi}(1 + \frac{1}{2\pi\beta}) \tag{6.12}$$

to a region where it is consistent within their large error with this result.

We used this prescription ($C = c_{opt}\beta$) and obtained estimates for the standard action starting from several values of β. Using this transformation the flow of Hamiltonians under blocking seems close to converging to a trajectory for many of the points simulated; all couplings come close to merging to a line. Thus the renormalized trajectory is probably close to the standard action, as seen in figure 6.4. However clear convergence is not obtained even for small couplings and as β increases some of the couplings do not get all the way to the curve. Thus the number of blocking steps required to reach

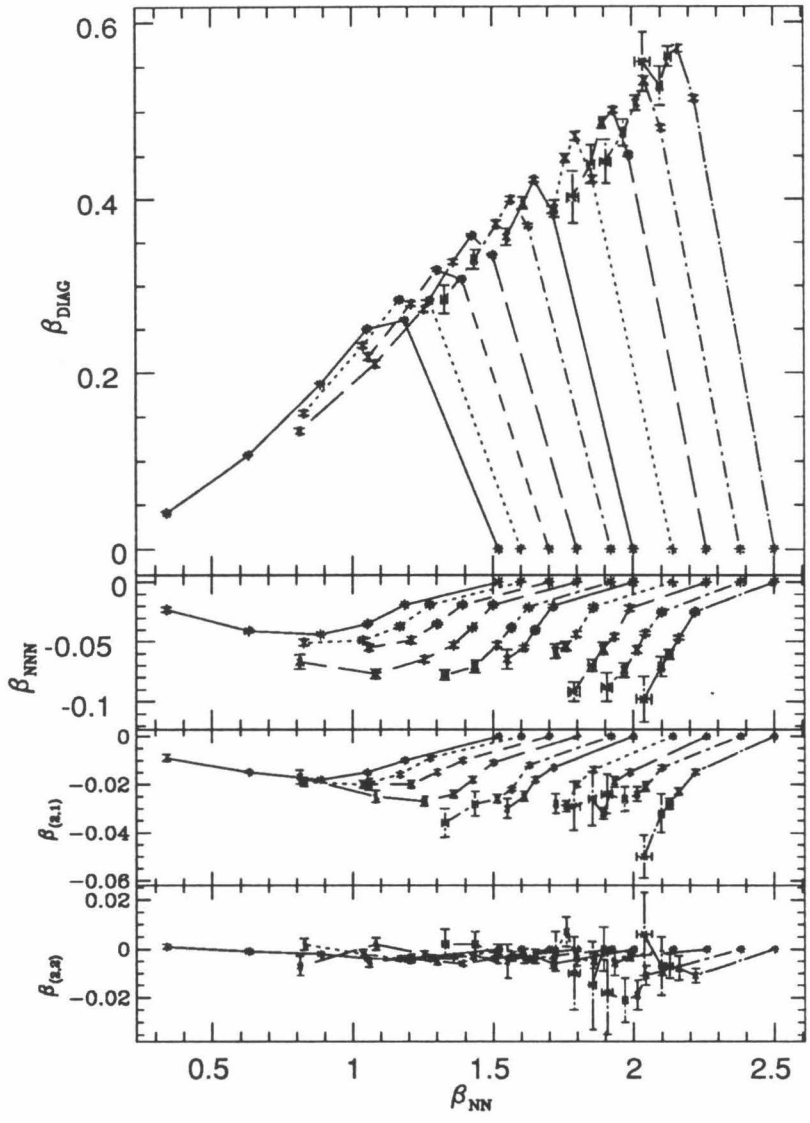

Figure 6.4: Flow of bilinear couplings under probabilistic block transformation with $C = c_{opt}\beta$, $c_{opt} = 2.3$. Convergence to a trajectory is not seen.

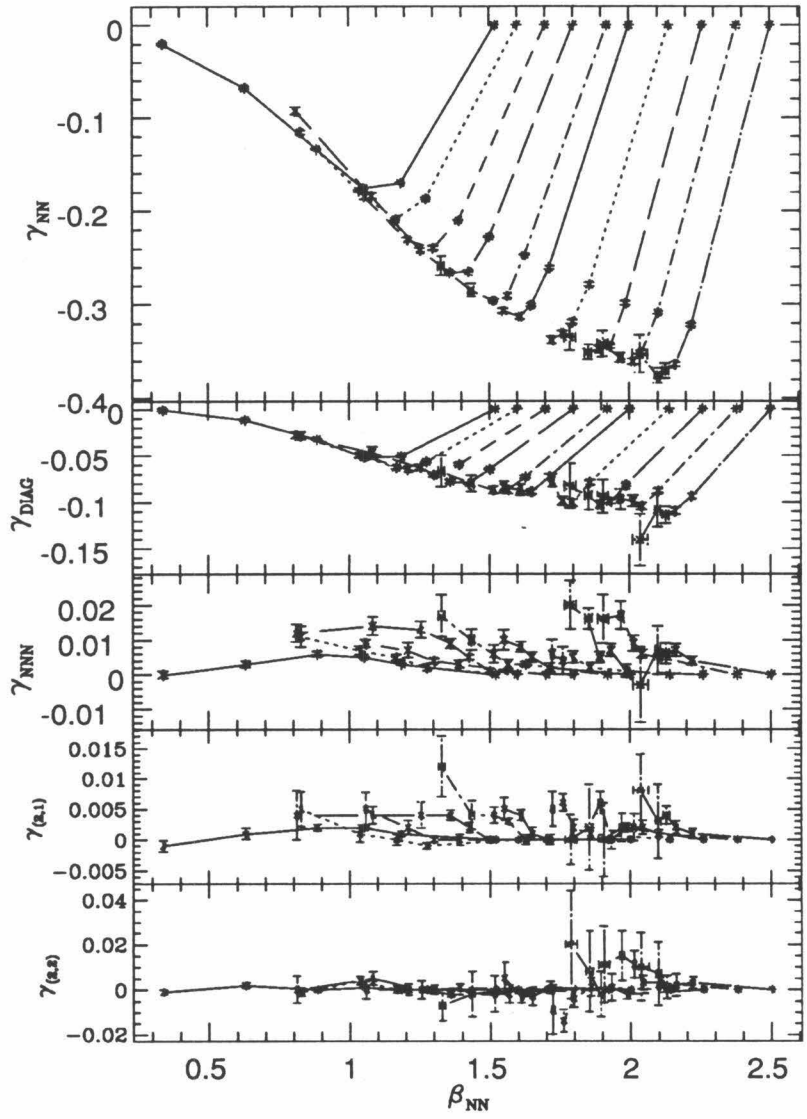

Figure 6.5: Flow of quartic couplings γ under probabilistic block transformation with $C = c_{opt}\beta$, $c_{opt} = 2.3$.

the trajectory gets large and it is not reached even after, for example, five blocking steps at $\beta = 2.0$.

In order to get around this, and in an attempt to find a transformation which converges to a renormalized trajectory in a few steps, we then tried the prescription $C = 5.0$ used by Hasenfratz and Niedermayer[19]. With it we obtained convergence to a renormalized trajectory within a few blocking iterations for all the couplings we measure - within the statistical errors of the measurements. In particular, as figures 6.6 and 6.7 show, proximity to the trajectory is obtained within two iterations and the points seem to lie on it after that.

	levels:		4/3		5/4		6/5
β	L	χ^2	$\Delta\beta$	χ^2	$\Delta\beta$	χ^2	$\Delta\beta$
2.14	512	129.3	0.1145(11)	5.2	0.1140(14)		
2.26	512	50.4	0.1177(20)	3.1	0.1149(9)		
2.38	512	71.9	0.1185(31)	9.4	0.1157(15)		
2.50	512	62.3	0.1236(20)	12.9	0.1163(16)		
2.50	1024	125.7	0.1200(2)	21.8	0.1183(18)	8.0	0.1166(36)
2.62	1024	137.4	0.1264(33)	4.5	0.1225(13)	10.2	0.1232(23)

Table 6.4: The discrete beta function $\Delta\beta(\beta)$ for the Standard Action from measurements using the Improved block transformation with $C = 5.0$ Shown are the coupling β and length L of the larger lattice, χ^2 and the value of $\Delta\beta$ at each the pair of levels matched.

Since such clear convergence to a trajectory is observed, we use the estimates for the coupling constants of the block Hamiltonians to obtain the

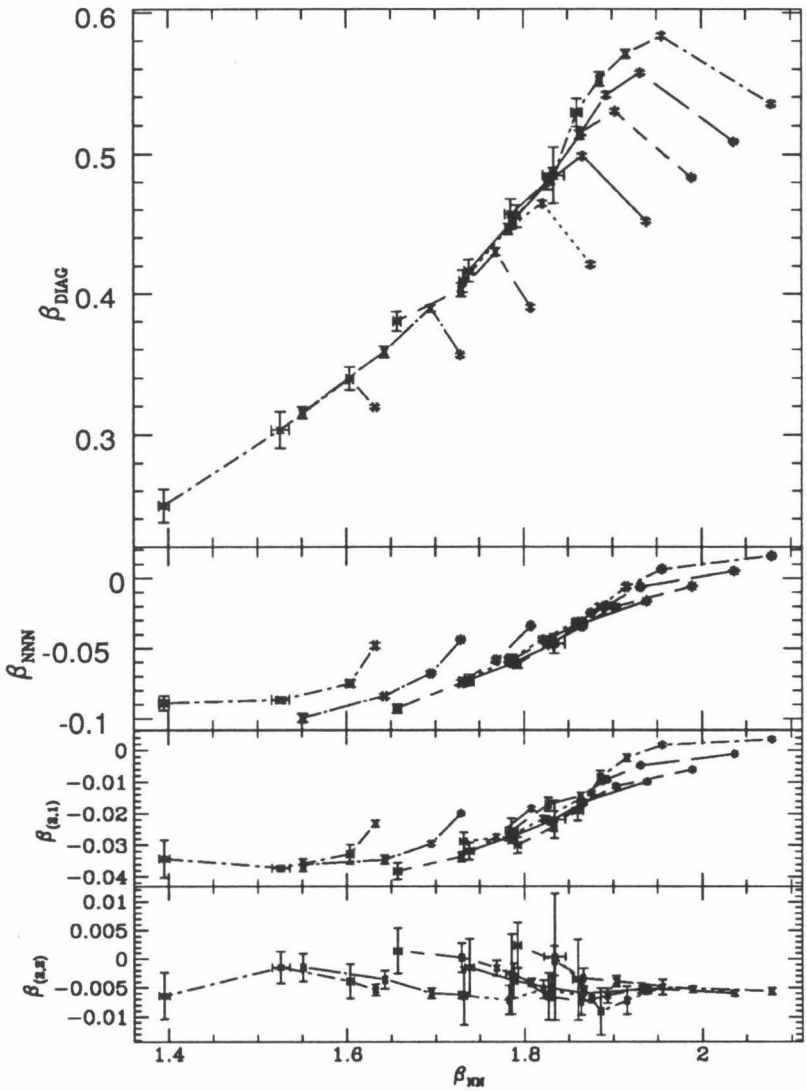

Figure 6.6: Flow of bilinear couplings under probabilistic block transformation with C=5.0 . Convergence to a trajectory is clear.

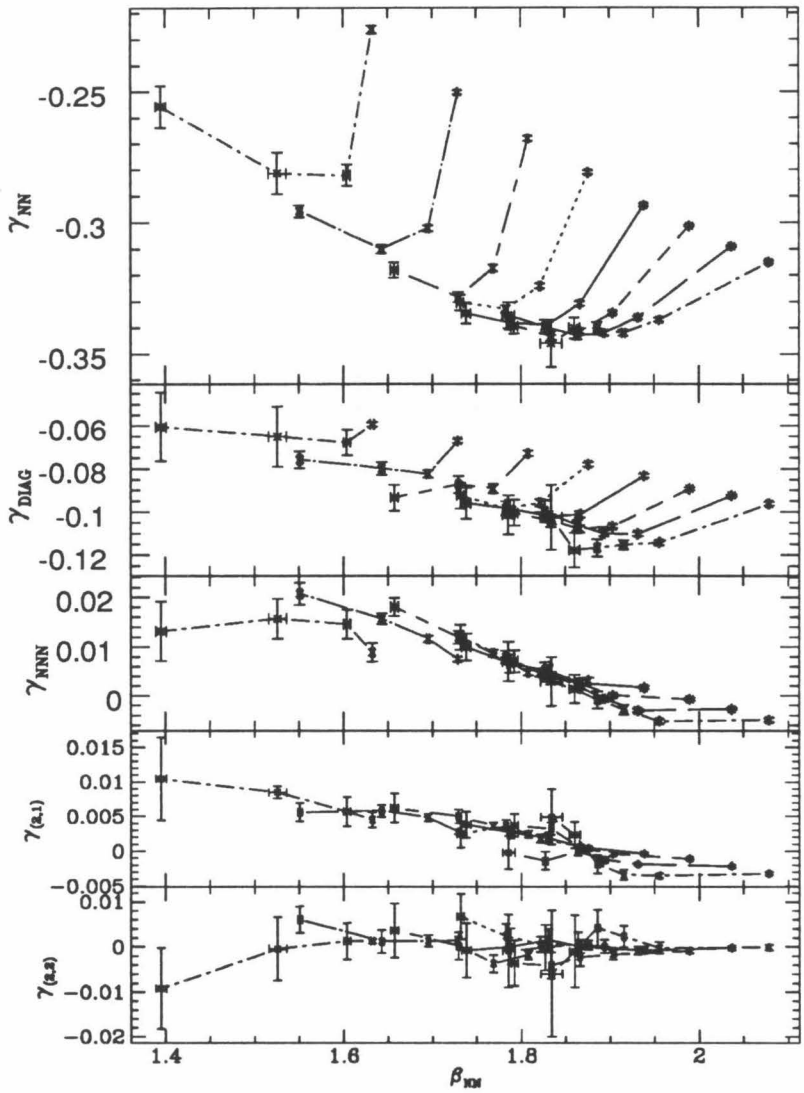

Figure 6.7: Flow of γ couplings under probabilistic block transformation with C=5.0 .

discrete beta function $\Delta\beta(\beta)$. To do this we fit the vector of couplings at each level to those obtained at one level less, on a lattice half the size, at several nearby coupling constant values. Intermediate values for the small lattice, between those simulated, are obtained by interpolation. We include in the fits the measured covariance matrix of the coupling constants obtained from the simulations. The estimates we get for $\Delta\beta(\beta)$ can be seen in table 6.4.

We note that the reason we include the covariance of the couplings is that there is a significant correlation between the different coupling constant values obtained (which is expected). In particular a check of a few typical examples shows that each coupling β_a and the corresponding γ_a are anti-correlated with linear correlation coefficients ranging between -0.55 and -0.98, while other coefficients range between -0.7 and 0.7.

Agreement between the results for subsequent levels is good in most of these cases, except for $\beta = 2.50$ at $L = 512$ to 256. However, since we are comparing actual coupling values, it is not as necessary to rely only on this agreement to check consistency - as is required when only a few correlation functions are used. We can also evaluate the values of χ^2 obtained in order to get a check of whether matching is good, and for the case quoted above we also have the match at larger lattice sizes which indicates good agreement.

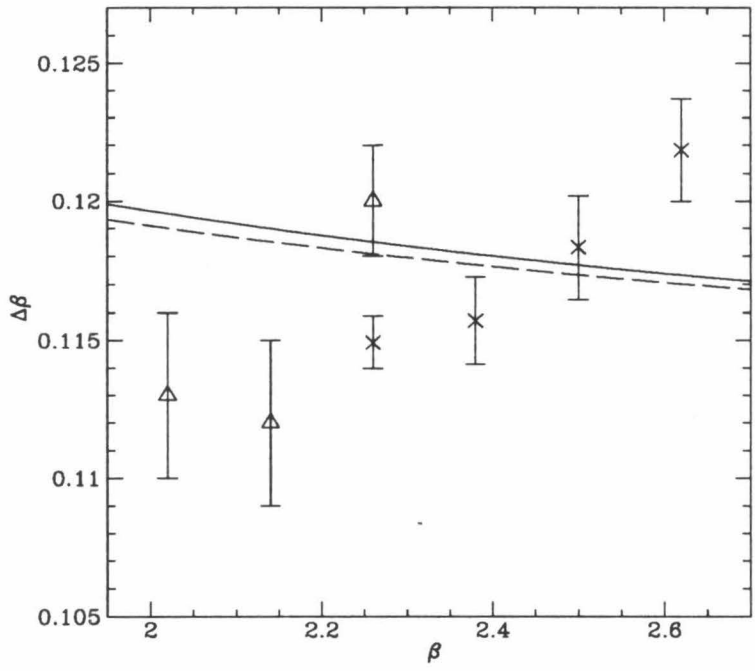

Figure 6.8: The discreet β function $\Delta\beta(\beta)$. Crosses are our measurements, triangles those of Hasenfratz *et al.* The solid line is the 3-loop perturbation theory result and the dashed line is the universal 2-loop result.

In general we see that matching the fourth level of blocking with the third (4/3) we get large χ^2, as the number of degrees of freedom is 9 (10 couplings minus one fit parameter). However matching deeper levels, 5/4 and where available 6/5, we see good values of χ^2. Since the fits at the 6/5 levels have much larger statistical errors we use the 5/4 match values as estimates for the discrete beta function $\Delta\beta(\beta)$.

A comparison of $\Delta\beta(\beta)$ with the results of Hasenfratz and Niedermayer[19] reveals that our result for the discrete beta function at $\beta = 2.14$ is in good agreement, while the one at $\beta = 2.26$ is lower and does not agree at about the 2.0 to 2.4σ level. Our results do not support their conclusion of convergence to the expected 3-loop result

$$\Delta\beta(\beta) = \ln 2/2\pi \left(1 + \frac{1}{2\pi\beta} + \frac{\kappa}{(2\pi\beta)^2} + O\left(\beta^{-3}\right)\right) \tag{6.13}$$

at $\beta = 2.26$. For the standard action $\kappa = 0.777$. As can be seen in figure 6.8 we observe slower movement towards the perturbative result, and only for $\beta = 2.38$ and 2.50 are our results consistent with it. The estimate at the largest β value, 2.62, however, overshoots the 3-loop result continuing the upward trend of the previous four points. This is a surprising result, since it is expected that any non-perturbative effects that would cause the deviations from the perturbative result would decay away in this regime of correlation lengths of a few thousand. The small change between the universal 2-loop and the standard action 3-loop result and the very slow variation of the additional term makes it unlikely that any single higher perturbative term could explain this large deviation. Thus if this estimate holds up it would be hard to reconcile with expectations.

A calculation of the next, 4-loop, term of the β function would nevertheless be very useful, as it would allow a check to see whether there are terms of oscillating sign that could contribute to the unusual behavior of the β function. It would be a difficult calculation, but as Falcioni and Treves[6] showed, it would require only a 3-loop calculation to obtain the difference of the 4th loop term for the standard lattice action from that for the modified minimal subtraction scheme. The latter has been calculated by Wegner[100].

We can use the measurements of the discrete beta function to obtain the value of the correlation length by making use of the fact that $\xi(\beta) = 2\xi(\beta - \Delta\beta(\beta))$. The results allow us to check the value of $m/\Lambda_{\overline{MS}}$, as a function of the coupling β, against the predicted exact value and against that measured directly in chapter 4. Checking for asymptotic scaling (using the coupling β) yields no new information, as it is equivalent to scaling for the discrete β function. However we can use it to check whether rescaling in the alternative coupling β_E improves the outlook for asymptotic scaling.

In figure 6.9 we show the ratio $m/\Lambda_{\overline{MS}}$of the mass gap to the lambda parameter in the modified minimal subtractions scheme obtained from this MCRG calculation and from direct Monte Carlo simulation on large lattices (chapter 4). The values shown include the two and three-loop estimates for the original coupling β and the rescaled coupling β_E that is derived from

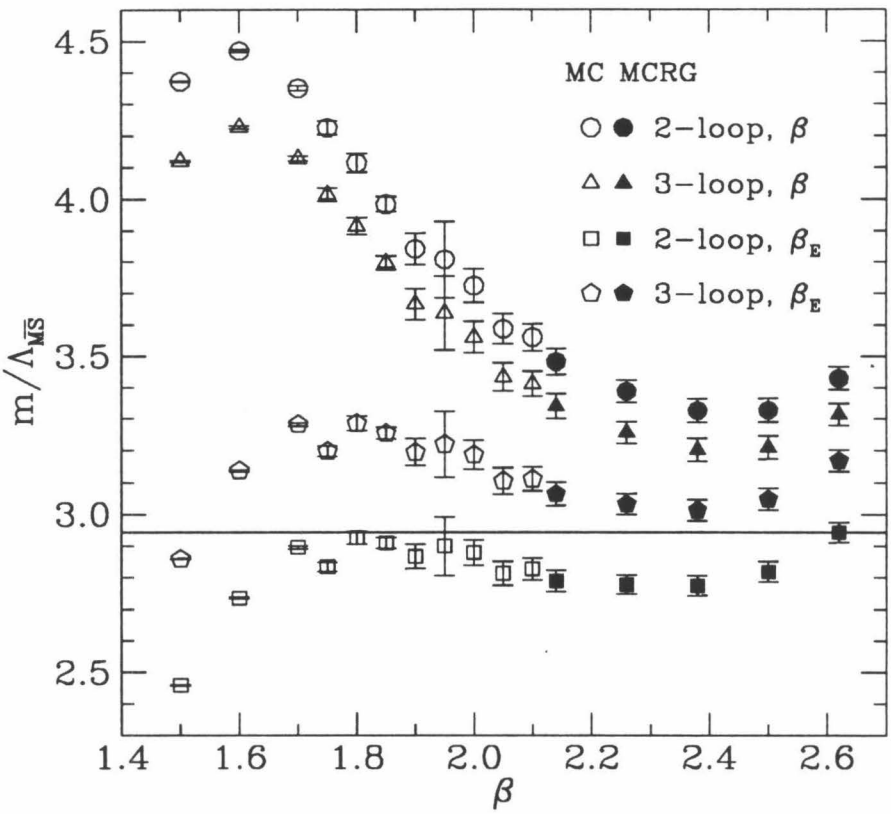

Figure 6.9: The mass gap to Λ parameter ratio for different values of the coupling constant β. The two-loop and three-loop results are shown, using the original coupling β and the rescaled one β_E, derived from the energy. Open points are from direct Monte Carlo measurements of the correlation length (chapter 4), and solid points from this MCRG calculation. The solid horizontal line is the exact result.

the energy. The first point of the MCRG data uses the correlation function estimate only at $\beta = 2.02$ (derived from the measurements of $m/\Lambda_{\overline{MS}}$at $\beta = 2.0$ and $\beta = 2.05$). We notice that the MCRG estimates at $\beta = 2.14$ and 2.26 and the values obtained by direct Monte Carlo measurements join smoothly.

We see that the value of $m/\Lambda_{\overline{MS}}$for the original coupling β remains more than 10% larger than the exact result $(m/\Lambda_{\overline{MS}})_{exact} = 2.94$ at the largest coupling measured: $m/\Lambda_{\overline{MS}}(\beta = 2.62) = 3.31(4)$. However the results using the two- or three-loop equations in the rescaled coupling constant β_E are compatible with asymptotic scaling between $\beta = 2.14$ and 2.50, and give a three-loop value of

$$m/\Lambda_{\overline{MS}}(\beta_E) = 3.047 \pm .035$$

at $\beta = 2.5$ - not far from the exact result. This tentative agreement is spoiled by the disagreement of the estimate at $\beta = 2.62$, $m/\Lambda_{\overline{MS}}(\beta_E) = 3.169(34)$; this effect is the same that made the measurement differ widely from the possible convergence of the discrete β function to the asymptotic result for smaller β.

Thus it would very interesting to obtain a check of our results using another method, in order to test the possible asymptotic scaling in the rescaled coupling β_E and see whether it extends to higher β values or whether the

discrete β function does actually show further non-monotonic behavior as our results at $\beta = 2.62$ suggest.

6.7 Conclusions and discussion.

We have demonstrated the utility of an interesting MCRG method coupled with the cluster MC update algorithm. We have confirmed the position of the renormalized trajectory for the standard block transformation, and showed how it extends to higher coupling constants. We have also shown that the probabilistic block transformation with 'optimal' constant choice $C = c_{opt}\beta$, with $c_{opt} = 2.3$ does not give good convergence to a renormalized trajectory when starting from the standard, nearest neighbor action.

We have obtained the renormalized trajectory for the same probabilistic block transformation with the choice $C = 5.0$ and used matching to obtain the discrete beta function of the theory. Our values for this are more accurate than those obtained by other methods and favor a slower convergence to the weak coupling series value than obtained by Hasenfratz *et al.* [19] (at $\beta = 2.24$) but after this agreement with the two-loop results our last point (at $\beta = 2.62$) overshoots it.

The results of Kim[101], making a new assumption for Finite Size Scal-

ing (FSS), indicate that the correlation length scales asymptotically for $\beta \geq 2.25$ but the susceptibility does not. In particular, assuming asymptotic scaling and that their FSS equation holds, they observe agreement with these assumptions for runs at $\beta = 2.246, 2.381, 2.509$ and 2.628. Our numbers are compatible with the two-loop discreet β function at 2.38 and 2.50 but not at 2.62; thus we agree with compatibility between the first three points of theirs, but not the last. Since they use the comparison to validate both their FSS method for this model (which is shown to work for the 2d Ising model) and asymptotic scaling, a fortuitous coincidence could hide discrepancies in both.

The unusual behavior of the long range physical quantities, the susceptibility and correlation length, has prompted a search for an explanation. As we have mentioned, Butera, Comi and Marchesini[23] used the Schwinger-Dyson equations to obtain high temperature expansions for the $O(N)$ model in two (and other) dimensions for general N through β^{11} and later extended them[102] to β^{14} using the series results of Luscher and Weisz[103] obtained using the linked cluster expansion . They saw that the critical singularity of the Ising and XY model exists on the real axis for $N \leq 2.4$ and then collides with another singularity and becomes a pair of singularities in the complex β plane. For small values of $N > 3$ these singularities are near the

real axis, and thus can influence the behavior of physical quantities, and for larger N they move further away. For $N = 3$ they estimate that they are at $\beta/N = \pm 0.63(6) \pm 0.12(6)i$. This would mean that, if the singularities were at the high end of the estimate $\beta = 1.9(2) \pm i0.4(2)$, they might be able to influence the behavior of physical quantities in the vicinity of our measurements, and be the cause of the large deviation from asymptotic scaling seen in this region.

Further study of these singularites is needed to resolve their location. This might be possible using an alternative method, the Monte Carlo approach of Marinari[104]. (Such a study is currently being tried by K. Anagnostopoulos using our simulation program). In principle this would also allow us to check whether singularities exist for larger values of the couplings, *i.e.*, $\beta > 2.4$. Furthermore it would provide us with a way of correcting for their presence, similar to that used by Bonnier and Hontebeyrie[47].

Lending support to this conclusion is the Monte Carlo simulation study by Klomfass *et al.* [105] that interpolates between the O(3) and O(2) (or XY) model. In this a mass term is added that couples to one component of the spin $\Delta H = +\gamma \sum_x (s_x^N)^2$ that tends to suppress that component. They attempt to explain the non-monotonic behavior of the β function around $\beta = 1.6$ by studying how the Kosterlitz-Thouless transition of the $O(2)$ model (a.k.a.

the XY model) extends in the γ, β plain from the KT point to the region near the O(3) model. They see that even for $\gamma = 1/15$ there is a transition at $\beta = 1.68(4)$ between a phase of bound vortices and a disordered phase like the only phase of the O(3) model. Thus they ascertain that the effect is a lattice specific, non-perturbative phenomenon which would thus get more pronounced the smaller that N gets (for $N > 3$), as, e.g., observed in Monte Carlo simulation by Wolff[22], where the $O(4)$ and $O(8)$ models are seen to exhibit smaller deviations from asymptotic scaling.

Our results show that when a rescaled coupling constant β_E is used, several values of $m/\Lambda_{\overline{MS}}(\beta_E)$ show agreement to two or three loops, giving the estimate $m/\Lambda_{\overline{MS}}(\beta_E) = 3.047(35)$ at three-loops. The disagreement of the last value, at $\beta = 2.62$ and the lack of a confirming calculation remain the major obstacles to firmly establishing this result.

We also show the utility of an interesting Monte Carlo Renormalization Group method that calculates the couplings of blocked Hamiltonians using the Schwinger Dyson equations. This has extracted the renormalized trajectories for two block transformations and provided very competitive estimates of the β function of the model.

Appendix A

Calculation of the 3-loop coefficient for β_E

We saw that using the rescaled temperature $\beta_E = \frac{N-1}{4(1-E)}$ and correcting for the universal two-loop behavior of the mass gap we obtain results much closer to asymptotic scaling than using the coupling β. Since the next term, the third-loop term, is non-universal and can provide an important correction to the leading behavior when we compare our results with the asymptotic formula for the mass gap we have undertaken to calculate it.

The three-loop term for the mass gap and susceptibility has been calculated for the modified minimal subtraction scheme of the continuum theory[5] and for several lattice actions[6]. In order to obtain the three-loop correction

for the mass gap for β_E, we will relate it to the value for the standard, nearest neighbor, lattice action.

For ease of calculation we will use the inverse of β_E, and we will call it $T_E = \frac{1}{\beta_E}$. To calculate the third-loop term for the mass gap we must get the same term for the β-function of the theory

$$\beta(T_E) = \frac{dT_E}{d(\log a)}. \tag{A.1}$$

From this definition it is obvious that we can obtain the rescaled values using the chain rule

$$\beta(T_E) = \frac{dT}{d(\log a)}\left(T(T_E)\right) \cdot \frac{dT_E}{dT}\left(T(T_E)\right) \tag{A.2}$$

and substituting T as a function of T_E into the resulting equation.

We proceed to calculate each part. From the weak coupling expansion of the energy for the standard action

$$E = 1 - \frac{N-1}{4}T - \frac{N-1}{32}T^2 - \frac{3(N-1)}{2000}\{5 + 4(N-1)\} \tag{A.3}$$

and the definition of T_E we can obtain a Taylor expansion for T_E as a function

of T

$$T_E = T + \frac{1}{8}T^2 + \frac{3\,[5 + 4(N-1)]}{500}T^3 + O(T^4). \tag{A.4}$$

We invert this and get

$$T = T_E - \frac{1}{8}T_E^2 + \frac{3\,[5 - 96(N-1)]}{4000}T_E^3 + O(T_E^4). \tag{A.5}$$

By using the perturbation expansion for the β-function of the standard lattice action

$$\beta(T) = T^2\left[b_0 + b_1 T + b_2 T^2 + O(T^3)\right], \tag{A.6}$$

where $b_0 = -\frac{1}{2\pi}$, $b_1 = -\frac{1}{(2\pi)^2}$ and $b_2 = -0.575/(2\pi)^3$, and appropriate substitutions we can obtain the same expansion for T_E,

$$\beta(T_E) = T_E^2\left[b_0 + b_1 T_E + h_2 T_E^2 + O(T_E^3)\right], \tag{A.7}$$

where h_2 is

$$h_2 = b_2 + \cdot\frac{1}{8}b_1 + \cdot\frac{192N - 77}{8000}b_0 \tag{A.8}$$

As it is easy to see that h_2 is simply related to the third-loop term of the expansion of the mass gap f_2, we can extract the difference in f_2 because of

$h_2 \neq b_2$, (sustituting $N = 3$)

$$\Delta f_2 = -\frac{1}{b_0^2}(h_2 - b_2) = -\frac{1}{8}\left\{1 - \frac{499}{500}\pi\right\} \tag{A.9}$$

Appendix B

Proof of the Schwinger-Dyson equations

In order to show that our formulation of the Schwinger-Dyson equations of the model is correct we will consider a correlation function $A(\{s\}) \equiv A$ that is invariant under $O(N)$ transformations of the spins. For ease of notation we will use $S = -H$ in our action. Then the integral

$$\int \prod_x d\mu(x)\, e^S\, A$$

will be invariant under infinitesimal local transformations of the spins

$$\sigma(x) \rightarrow \sigma'(x) = \left(1 + \sum_{\alpha} \epsilon^{\alpha}(x) T^{\alpha}\right) \sigma(x)$$

where T^{α} are the $\frac{1}{2}N(N-1)$ generators of O(N) rotations. This leads to the condition

$$\int \prod_{x} d\mu(x) \left[\exp\left(\sum_{\alpha} \epsilon^{\alpha}(x) E^{\alpha}(x)\right) - 1\right] e^{S} A = 0 \qquad \text{(B.1)}$$

where $E^{\alpha}(y)$ are the local differential operators that generate rotations at site y and obey the condition $[E^{\alpha}(y), \sigma(x)] = \delta_{xy} T^{\alpha} \sigma(x)$. Since $O(N)$ invariance means that averages of odd powers of the fields are zero, taking the first term in the Taylor expansion in ϵ gives a trivial equation. The second term in the expansion yields

$$\int \prod_{x} d\mu(x) E^{\alpha}(y) E^{\beta}(z) e^{S} A = 0 \qquad \text{(B.2)}$$

Furthermore, because for $y \neq z$ or $\alpha \neq \beta$ the resulting quantities are not O(N) invariant, and must be zero, the terms left are those for which $y = z$ and $\alpha = \beta$. From that we obtain

$$\int \prod_{x} d\mu(x) \sum_{\alpha} E^{\alpha}(y) e^{S} \left\{A E^{\alpha}(y) S + E^{\alpha}(y) A\right\} = 0 \qquad \text{(B.3)}$$

Because the domain is compact and the measure is invariant the two terms each must be seperately zero. Thus by taking the right part and dividing by the partition function Z we obtain an equality of expectation values, *i.e.*,

$$\left\langle \sum_{\alpha} \left(E^{\alpha}(y)S\right)\left(E^{\alpha}(y)A\right) \right\rangle + \left\langle \sum_{\alpha} E^{a}(y)E^{a}(y)A \right\rangle = 0. \qquad \text{(B.4)}$$

Up to here we followed the derivation of Butera *et al.* [23]. Now we use the definition of $E^{a}(y)$ to obtain $[E^{\alpha}(y), B] = \frac{\partial B}{\partial \sigma(y)} T^{\alpha}\sigma(y)$ and substitute this in the previous equation. We also drop the $< \ldots >$ for convenience and take equations to mean an equality of expectation values. We get

$$\begin{aligned} &\sum_{\alpha} \left(\frac{\partial S}{\partial \sigma(y)} T^{\alpha}\sigma(y)\right)\left(\frac{\partial A}{\partial \sigma(y)} T^{\alpha}\sigma(y)\right) + \\ &+ \sum_{\alpha}\sum_{ik} \frac{\partial}{\partial \sigma^{i}(y)}\left(\frac{\partial A}{\partial \sigma^{l}(y)}\right) T^{\alpha}_{ik}\sigma^{k}(y) \sum_{lm} T^{\alpha}_{lm}\sigma^{m}(y) + \\ &+ \sum_{\alpha}\sum_{lmn} \frac{\partial A}{\partial \sigma^{i}(y)} T^{\alpha}_{lm} T^{\alpha}_{mn}\sigma^{n}(y) = 0. \end{aligned}$$

We then use the properties of the generators T^{α}, namely

$$\sum_{\alpha} T^{\alpha} T^{\alpha}_{kl} = \delta_{il}\delta_{jk} - \delta_{ik}\delta_{jl}$$

and we obtain

$$S^{a}_{,y}\sigma(y)^{a}\sigma(y)^{b}A^{b}_{,y} - S^{a}_{,y}A^{a}_{,y} + A^{a\ b}_{,y,y}\sigma(y)^{a}\sigma(y)^{b} - A^{a\ a}_{,y,y} + (N-1)A^{a}_{,y}\sigma(y)^{a} = 0.$$

This can be more compactly expressed using the projection operator $P^{ab}(y)$ as

$$-A^{a}_{,i}P^{ab}_{i}S^{b}_{,i} - P^{cd}_{i}A^{cd}_{,ii} + (N-1)\sigma^{a}_{i}A^{a}_{,i} = 0.$$

We get our equation of motion if we also substitute $S = -H$ and use the definition of the generalized force $F^{b}_{i} = -H^{b}_{,i}$:

$$A^{a}_{,i}P^{ab}_{i}F^{b}_{i} + P^{cd}_{i}A^{cd}_{,ii} = (N-1)\sigma^{a}_{i}A^{a}_{,i}.$$

Bibliography

[1] A.M. Polyakov, *Phys. Lett.* **59B**, 79 (1975) .

[2] C.J. Hamer, J.B. Kogut, L. Susskind, *Phys. Rev.* **D19**, 3091 (1979).

[3] S. H. Shenker and J. Tobochnik, *Phys. Rev.* **B22**, 4462 (1980).

[4] E. Brezin and J. Zinn-Justin, *Phys. Rev.* **B14** 3110 (1976) .

[5] S. Hikami and E.Brezin, *J. Phys.* **A 11**, 1141 (1978).

[6] M.Falcioni and A.Treves, *Phys. Lett.* **159B**, 140 (1985); *Nucl. Phys.* **B265**, 671 (1986).

[7] G. Parisi, *Phys. Lett.* **92B**, 133 (1980).

[8] G. Martinelli, G. Parisi and R. Petronzio, *Phys. Lett.* **100B**, 485 (1981).

[9] B.Berg and M.Luscher, *Nucl. Phys.* **B190** [FS3] 412 (1981).

[10] G. Fox, R. Gupta, O. Martin and S. Otto, *Nucl. Phys.* **B205**, 188 (1982).

[11] A. Hasenfratz and A. Margaritis, *Phys. Lett.* **133B** 211 (1983); *ibid* **148B** 129 (1984).

[12] B. Berg, S. Meyer and I. Montvay, *Nucl. Phys.* **B235** [FS11], 149 (1984).

[13] U. Wolff, *Phys. Lett.* **222B**, 473 (1989) and *Nucl. Phys.* **B334**, 581 (1990).

[14] J. Apostolakis, C. Baillie and G.C. Fox, *Phys. Rev.* **D43**, 2687 (1991).

[15] Alexander B. Zamolodchikov and Alexsei B. Zamolodchikov, *JETP Lett.* **26**, 457 (1977)

[16] P.B.Wiegmann, *Phys. Lett.* **152B**, 209 (1985).

[17] A.M.Polyakov and P.B.Wiegmann, *Phys. Lett.* **131B**, 121 (1983).

[18] P. Hasenfratz, M. Maggiore and F. Niedermayer, *Phys. Lett.* **245B**, 522 (1990)

[19] P. Hasenfratz and F. Niedermayer, *Nucl. Phys.* **B337**, 233 (1990).

[20] H. Flyvbjerg, F. Larsen and C. Kristjansen, in 'Lattice 90', proceedings of the International Symposium, Tallahasee, Florida 1990, *Nucl. Phys. B (Proc. Suppl.)*, 44 (1991).

[21] S.Samuel, O.Martin and K.Moriarty, *Phys. Lett.* **153B**, 87 (1985).

[22] U.Wolff, *Phys. Lett.* **248B**, 335 (1990).

[23] P. Butera, M.Comi and G.Marchesini, *Nucl. Phys.* **B300**, 1 (1988).

[24] S.L.Adler, *Phys. Rev.* **D37**, 458 (1988); H.Neuberger, *Phys. Rev. Lett.* **59**, 1877 (1987).

[25] R. Gupta, J. DeLapp, G. G. Batrouni, G. C. Fox, C. F. Baillie and J. Apostolakis, *Phys. Rev. Lett.* **61**, 1996 (1988).

[26] U. Wolff, *Phys. Lett.* **288B**, 166 (1992).

[27] G. Bathas and H. Neuberger, *Phys. Rev.* **D45**, 3880 (1992).

[28] R.H. Swendsen and J.S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987).

[29] Wolff U., *Phys. Rev. Lett.* **62**, 361 (1989).

[30] A. Hasenfratz, P. Hasenfratz, U. Heller and F. Karsch, *Phys. Lett.* **140B**, 76 (1984) and *Phys. Lett.* **143B**, 193 (1984).

[31] N.D. Mermin and H. Wagner, *Phys. Rev. Lett.* **22**, 1133 (1966); N.D. Mermin, *Jour. of Math. Phys.* **8**, 1061 (1967).

[32] S.Coleman, *Commun. Math. Phys* **31**, 259 (1973).

[33] J.M.Kosterlitz, D.J.Thouless, *Jour. of Phys.* **C6**, 1181 (1973).

[34] A.McKane and M.Stone, *Nucl. Phys.* **B163**, 169 (1980).

[35] D.J. Amit and G. Kotliar, *Nucl. Phys.* **B170** [FS1], 187 (1980).

[36] S. Elitzur, *Nucl. Phys.* **B212**, 501 (1983); A.Jevicki, *Phys. Lett.* **61B**, 327 (1977).

[37] K. Pohlmeyer, *Comm. Mathem. Phys.* **46**, 207 (1976); A.M. Polyakov, *Phys. Lett.* **72B**, 224 (1977); M. Luscher, *Nucl. Phys.* **B135**, 1 (1978).

[38] A.B.Zamolodchikov and A.B.Zamolodchikov, *Nucl. Phys.* **B133**, 525 (1978).

[39] B.Berg, M.Karowski, V.Kurak and P.Weisz, *Phys. Lett.* **76B**, 502 (1978).

[40] K.Binder and D.W.Heermann, "Monte Carlo Simulation in Statistical Physics", Spring-Verlag, Berlin Heidelberg 1988.

[41] A.D.Sokal, "Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms," Lecture in "Troisieme Cycle de la Physique en Suisse Romande" (1989).

[42] N. Metropolis *et al.*, *J. Chem. Phys.* **21**, 1087 (1953).

[43] W.K.Hastings, Biometrica **57**, 109 (1970).

[44] S.L.Adler, *Phys. Rev.* **D23** 2901 (1981).

[45] P.W.Kasteleyn and C.M.Fortuin, *J. Phys. Soc. Jpn. Suppl.* **26s**, 11 (1969); C.M.Fortuin and P.W.Kasteleyn, *Physica (Utrecht)* **57**, 536 (1972).

[46] R.G.Edwards and A.D.Sokal, *Phys. Rev.* **D40**, 1374 (1989).

[47] B.Bonnier and M.Hontebeyrie, *Phys. Lett.* **226B**, 361 (1989).

[48] M.Luscher and P.Weiss, *Nucl. Phys.* **B300**, 325 (1988).

[49] P. Hasenfratz and F. Niedermayer, *Phys. Lett.* **245B**, 529 (1990).

[50] J-M Drouffe and H. Flyvbjerg, *Phys. Lett.* **206B** 285 (1988) and *Nucl. Phys.* **BB332** 687 (1990); H. Flyvbjerg, *Phys. Lett.* **219B**, 323 (1989).

[51] H. Flyvbjerg, in 'Lattice 89', proceedings of the International Symposium, Capri, Italy, 1989, edited by R. Petronzio *et al.* Nucl. Phys. B (Proc. Suppl.) **17** 343 (1990); *Nucl. Phys.* **B344**, 646 (1990).

[52] M.Hasenbusch, S.Meyer and G.Mack, in 'Lattice 90', proceedings of the International Symposium, Tallahasee, Florida 1990, edited by U.M. Heller, A.D. Kennedy and S.Sanielevici *Nucl. Phys. B (Proc. Suppl.)* **20**, 110 (1991).

[53] M.Luscher, P.Weisz and U.Wolff, *Nucl. Phys.* **B359**, 221 (1991).

[54] M.Creutz, *Phys. Rev.* **D36** 515 (1987).

[55] F.R.Brown and T.W.Woch, *Phys. Rev. Lett.* **58** 2394 (1987).

[56] See for example the appendices of: R. Gupta *et al.* , *Phys. Rev.* **D36**, 2813 and S. Gottlieb *et al.* , *Nucl. Phys.* **B263**, 704 (1986). For more details see B.Efron, *Ann. Stat.* **7**, 1 (1979).

[57] M. Luscher, in *Progress in Field Theory*, proceedings of the Cargese Summer Institute , Cargese, France, 1983, edited by G. 't Hooft, A. Singer, and R. Stora, NATO Advanced Study Institute, Serries B: Physics, Vol. 115 (Plenum, New York, 1984).

[58] I.Bender, W.Wetzel and B.Berg, *Nucl. Phys.* **B269**, 389 (1986).

[59] H.Neuberger, *Phys. Lett.* **207B**, 461 (1988).

[60] A.D.Sokal, in 'Lattice 90', Proceedings of the International Conference, Tallahassee, Florida, 1990, edited by U.M. Heller, A.D. Kennedy, and S. Sanielevici *Nucl. Phys. B (Proc. Suppl.)* **20**, 55 (1991).

[61] U.Heller and H.Neuberger, *Phys. Rev.* **D39**, 616 (1989).

[62] N.Madras and A.D.Sokal, *J. of Stat. Phys.* **50**, 179 (1988).

[63] This was pointed out to us by A.D. Sokal.

[64] M. Evans, *Nucl. Phys.* **B208**, 122 (1982).

[65] M. Klomfass, U.M.Heller and H. Flyvbjerg, *Nucl. Phys.* **B360**, 264 (1991).

[66] G. Parisi, *Statistical Field Theory,* (Addison-Wesley, Reading, Mass., 1988).

[67] *Monte Carlo Methods in Statistical Physics, Topics in Current Physics 7, Second Edition*, Ed. K. Binder (Springer-Verlag, Berlin, 1986); *Applications of the Monte Carlo Method in Statistical Physics, Topics in Current Physics 36, Second Edition*, Ed. K. Binder (Springer-Verlag, Berlin, 1987).

[68] A. D. Sokal, in *Computer Simulation Studies in Condensed Matter Physics: Recent Developments*, eds. D. P. Landau *et al.* (Springer-Verlag, Berlin, 1988).

[69] U. Wolff, in Proc. of the Symposium on Lattice Field Theory, Capri, September 1989, *Nucl. Phys. B (Proc. Suppl.)* **17**, 93 (1990).

[70] J.-S. Wang and R. H. Swendsen, *Physica A* **167**, 565 (1990).

[71] E. M. Reingold, J. Nievergelt and N. Deo, *Combinatorial Algorithms: Theory and Practice* (Prentice-Hall, Englewood Cliffs, N.J., 1977); E. Horowitz and S. Sahni, *Fundamentals of Computer Algorithms*, (Computer Science Press, Rockville, Maryland, 1978).

[72] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, (Academic Press, New York, 1982).

[73] C.F. Baillie and P.D. Coddington, Concurrency: Practice and Experience **3**, 129 (1991).

[74] R. G. Edwards, X.-J. Li and A. D. Sokal, *Sequential and Vectorized Algorithms for Computing the Connected Components of an Undirected Graph*, in preparation.

[75] D. Hillis, *The Connection Machine*, (MIT Press, Cambridge, Mass., 1985).

[76] R. C. Brower, P. Tamayo and B. York, *J. Stat. Phys.* **63**, 73 (1991).

[77] E. N. Miranda, *Physica A* **175**, 229 (1991).

[78] *Programming in Paris*, (Thinking Machines Corporation, Cambridge, Mass., 1989).

[79] J. Apostolakis, P. Coddington and E. Marinari, *Europhys. Lett.* **17**, 189 (1992).

[80] Talk presented at the conference "Physics Computing '91," San Jose, California, June 1991.

[81] W. D. Hillis and Guy L. Steele Jr., *Comm. of the ACM* **29**, 1170 (1986).

[82] B. A. Galler and M. J. Fisher, *Commun. ACM* **7**, 301 (1964); D. E. Knuth, *Fundamental Algorithms*, vol. 1 of *The Art of Computer Programming* (Addison-Wesley, Reading, Mass., 1968).

[83] J. Hoshen and R. Kopelman, *Phys. Rev. B* **14**, 3438 (1976).

[84] Y. Shiloach and U. Vishkin, *J. Algorithms* **3**, 57 (1982).

[85] D. Stauffer, *Introduction to Percolation Theory*, (Taylor and Francis, Philadelphia, 1985).

[86] P. Rossi and G. P. Tecchiolli, *Finding Clusters in a Parallel Environment*, unpublished.

[87] S.K. Ma, *Phys. Rev. Lett.* **37**, 461 (1976).

[88] R.H. Swendsen, *Phys. Rev. Lett.* **42**, 859 (1979).

[89] R.H. Swendsen, in *Real Space Renormalization, Topics in Current Physics*, Volume **30**, edited by Th.W. Burkhardt and J.M.J. van Leeuwen

[90] K.G. Wilson, in *Recent Developments in Gauge Theories*, edited by G. 't Hooft *et al.* (Plenum, New York 1980).

[91] J.E. Hirsch and S.H. Shenker, *Phys. Rev.* **B27**, 1736 (1983).

[92] R.H. Swendsen, *Phys. Rev. Lett.* **52**, 2321 (1984).

[93] R. Gupta, in *Lattice Gauge Theory – A challenge in large scale computing*, proceedings of NATO Workshop, Wuppertal, Germany, 1985, edited by B. Bunk, K.H. Mutter, and K. Schilling, (Plenum Press, New York, 1986).

[94] M. Falcioni, G. Martinelli, M.L. Paciello, G. Parisi, and B. Taglienti, *Nucl. Phys.* **B265**[FS15] 187 (1986).

[95] M. Creutz, *Phys. Rev. Lett.* **50**, 1411 (1983); D.J.E. Callaway and A. Rahman. *Phys. Rev. Lett.* **49**, 613 (1982).

[96] M. Creutz, A. Gocksch, M. Ogilvie, and M. Okawa, *Phys. Rev. Lett.* **53**, 857 (1984).

[97] R.H. Swendsen, *Phys. Rev. Lett.* **52**, 1165 (1984).

[98] P. Stolorz, *Phys. Lett.* **172B**, 77 (1986).

[99] M. Okawa, *Phys. Rev. Lett.* **54**, 963 (1985).

[100] F. Wegner, *Nucl. Phys.* **B316**, 663 (1989) and W. Bernreuther and F.J. Wegner, *Phys. Rev. Lett.* **57**, 1383 (1986).

[101] J-K Kim, *Phys. Rev. Lett.* **70**, 1735 (1993).

[102] P. Butera, M.Comi and G.Marchesini, *Phys. Rev.* **B41**, 11494 (1990).

[103] M.Luscher and P.Weisz, *Nucl. Phys.* **B300**, 325 (1988)

[104] E.Marinari, *Nucl. Phys.* **B235** [FS11], 123 (1984).

[105] M. Klomfass, U.M. Heller and H. Flyvbjerg, *Nucl. Phys.* **B360**, 264 (1991).