High Precision Simulations with Fast Algorithms ¹

Ulli Wolff

CERN, Theory Division,
CH-1211 Genève 23, Switzerland

Abstract

A brief non-rigorous introduction to the Monte Carlo method for the simulation of lattice quantum field theoretical models is given. The associated problem of critical slowing down in the continuum limit is addressed together with its partial solution for some classes of systems. The version of overrelaxation mostly used in practice, with alternating microcanonical and standard updates, is analyzed exactly for free fields and described in numerical applications to spin and gauge systems. Cluster algorithms are briefly introduced, and applications to the two dimensional $O(3)$ σ -model are described in some detail. They include the exploration of scaling behavior, a numerical estimate of scattering phases and the demonstration and test of a general technique to compute running renormalized coupling constants in asymptotically free theories. We conclude with remarks on the present status of algorithm development.

CERN-TH 6406/92
March 1992

¹Lecture notes for the 31. Internationalen Universitätswochen für Kern- und Teilchenphysik, Computational Methods in Field Theory, Schladming 1992

1 Introduction

Quantum field theories for the description of elementary particles and their interactions are only formally defined by specifying fields and an action with certain symmetries. The actual derivation of predictions requires a limiting procedure where one first regularizes the theory under study at short distance with an ultraviolet cutoff. This ad hoc modification is unphysical and presumably has to do with the end of the domain of validity of the field theoretical model. Only the cutoff independent long distance features of renormalizable theories can be hoped to approximate the underlying physics at present energies with all reference to shorter distances embodied in some phenomenological parameters.

One possible cutoff procedure is to discretize space time by a lattice with spacing a . It has the virtue of manifestly reducing the number of degrees of freedom per volume to become finite and makes thus a numerical computation straightforward in principle. Apart from the spacing one has the system size L and physical scales like masses m as dimensionful quantities. Finite size effects characterized by the dimensionless combination mL , although also artificial, are often not too difficult to control and sometimes even a useful way to probe the system. On the other hand $\xi = (ma)^{-1}$, which has to be extrapolated to infinity to remove the cutoff, is more problematic. It corresponds to a correlation length in units of lattice spacings, and when ξ gets large the problem of critical slowing down arises in Monte Carlo simulations. Depending on how quantities of interest approach their asymptotic value — proportional to ξ^{-2} or ξ^{-1} or only $1/\ln \xi$ — it becomes difficult to verify convergence and to extrapolate. It is in this area that significant progress has been made over the last five years, some of which will be reported in these lectures².

In some interesting two dimensional models (nonlinear σ -models) correlation lengths of order $\xi = 100$ can be simulated accurately now. Some appropriate physical quantities get close enough to the cutoff limit then to reliably extrapolate them to zero lattice spacing. The progress up to now has been less pronounced in four dimensional gauge theories. Cluster and multigrid techniques, which led to the most spectacular breakthrough for

²As the field is under very active study, this overview cannot be complete. The choice of examples presented here is biased toward the author's contributions. In particular, no small-step methods, as they are used in connection with fermions, are discussed.

σ -models, are not (yet) fully developed, and it is not yet possible to have as large systems as in two dimensions. However, with massively parallel computers and further algorithmic progress we may well arrive at a similar computing ability for four dimensions.

The following notes are divided into four parts. After this introduction we start with a simple non-rigorous introduction to Monte Carlo simulation and define parameters that allow us to judge the quality of an algorithm. In section 3 overrelaxation algorithms are discussed which, although outperformed in some cases, are at present the most universally applicable tool (including gauge theory). In section 4 the cluster method is reviewed. While the description of the algorithm is kept brief, most space is devoted to physical results in two dimensions that have been obtained. We conclude with an outlook on the present situation with regard to algorithm development.

2 Monte Carlo evaluation of path integrals

2.1 Free fields on the lattice

In this subsection we develop some notation by briefly discussing free fields on the lattice. They will furthermore be useful as an example, where also the dynamical properties of Monte Carlo algorithms can be derived analytically.

We consider a D -dimensional hypercubic lattice with sites

$$x = (x^0, x^1, \dots, x^{D-1}), \quad x^\mu = n^\mu a, \quad (1)$$

where the n^μ are integer and a is the dimensionful lattice spacing. We will mostly use lattice units, $a = 1$, but reintroduce a on some special occasions to emphasize its rôle for the continuum limit. A real scalar field $\varphi(x)$ is defined on each site, and, with periodic boundary conditions of equal period L (L/a integer) in all directions, it obeys

$$\varphi(x) = \varphi(x + L\hat{\mu}) \quad (2)$$

with $\hat{\mu}, |\hat{\mu}| = 1$, a unit vector along any of the axes. We may thus take $(L/a)^D$ field values as independent, for instance those with

$$\{\varphi(x) \mid 0 \leq x^\mu < L\}. \quad (3)$$

We further need the lattice derivative

$$\partial_\mu \varphi(x) = \frac{\varphi(x + \hat{\mu}a) - \varphi(x)}{a}, \quad (4)$$

and may now write the free field action

$$S(\varphi) = \frac{1}{2} \sum_x a^D \left\{ \sum_\mu (\partial_\mu \varphi(x))^2 + m^2 \varphi(x)^2 \right\} \quad (5)$$

which, for smooth classical fields sampled on a grid, tends to the integral $\frac{1}{2} \int d^D x \{(\partial_\mu \varphi)^2 + m^2 \varphi^2\}$ in the limit $a \rightarrow 0$ at fixed L .

For particle physics we wish to integrate over all sampled fields and extract the physics from (euclidean) Green functions of the type

$$\langle \varphi(x_1) \cdots \varphi(x_n) \rangle = \frac{1}{Z} \int \prod_x d\varphi(x) e^{-S(\varphi)} \varphi(x_1) \cdots \varphi(x_n), \quad (6)$$

where Z is the normalizing partition function. We would like to emphasize a few points about this formulation of a quantum field theory:

- It is approximated by a high but finite dimensional integral, which is well defined for $m^2 > 0$ in (5) and can in principle be computed numerically.
- This is still true when an interaction term like $\lambda \varphi^4$ is added to (5); the definition is independent of perturbation theory and Feynman diagrams.
- These nice properties are due to the presence of both an ultraviolet cutoff $a \neq 0$ and an infrared cutoff $L < \infty$, whose rôles have to be kept track of, for instance by extrapolating.

A simple and important observable allows measurement of the mass gap. The standard procedure is first to define fields at zero spatial momentum

$$\phi_0(t) = \sum_{\vec{x}} \varphi(x^0 = t, \vec{x}) \quad (7)$$

that depend on the coordinate $x^0 = t$ by convention singled out as euclidean time. The gap $\hat{m}$ follows from the observed temporal decay

$$\langle \phi_0(t) \phi_0(0) \rangle = \text{const} \left(e^{-t\hat{m}} + e^{-\hat{m}(L-t)} \right), \quad (8)$$

where the second term takes into account periodicity in time. In the free case the energy $\hat{m}$ is related to m by

$$m = \frac{2}{a} \sinh(\hat{m}a/2), \quad \hat{m} = m(1 - \frac{1}{24}(ma)^2 + \dots). \quad (9)$$

Here we have an instance of a (UV-)cutoff correction or lattice artifact. The expected relation $\hat{m} = m$ emerges only in the cutoff or continuum limit $a \rightarrow 0$ at fixed m , where the correlation length in lattice units $\xi = (\hat{m}a)^{-1}$ diverges. In other words, we tune the system to a critical point. The renormalized mass in the presence of interactions would still be extracted by fitting with (8) at large separation. Similar procedures allow us to derive renormalized couplings and other physical quantities from Green functions.

2.2 Monte Carlo simulation of lattice path integrals

The clue to a numerical evaluation of (6) is the observation that it is a structure known from equilibrium statistical mechanics. In

$$\langle A(\varphi) \rangle = \frac{1}{Z} \int \prod_x d\varphi(x) e^{-S(\varphi)} A(\varphi) \quad (10)$$

the exponential Boltzmann factor controls the support of the integral in phase space with the observable $A(\varphi)$ as a slowly varying modification. This is true at least for certain classes of observables like polynomial ones, and for some of them the Monte Carlo techniques can be hoped to be practical.

We generate a sequence of field configurations $\varphi^{(n)}$ in a computer memory,

$$\varphi^{(0)} \rightarrow \varphi^{(1)} \rightarrow \varphi^{(2)} \rightarrow \dots, \quad (11)$$

that have the Boltzmann part “built in”, i.e. for late members in the sequence (large enough n) configurations have a probability of appearance

$$P(\varphi^{(n)} = \varphi) = \frac{1}{Z} \exp(-S(\varphi)). \quad (12)$$

Then an approximation to (10) is given by the sample average

$$\langle A(\varphi) \rangle \approx \frac{1}{N} \sum_{i=n+1}^{i=n+N} A(\varphi^{(i)}). \quad (13)$$

In (11) the recursion starts from some initial $\varphi^{(0)}$, and $\varphi^{(i)} \rightarrow \varphi^{(i+1)}$ means that for given $\varphi^{(i)}$ we choose $\varphi^{(i+1)}$ at random with a probability $T(\varphi^{(i)}, \varphi^{(i+1)})$. The set of probabilities T together with a constructive prescription on how to realize the transition make up a Monte Carlo algorithm. There are clearly constraints from (12), and furthermore we have to address the questions of how large n has to be and what “ $\approx$ ” means in (13).

For some arguments to follow now it is helpful to think of an ensemble of independent simulations of the kind just discussed. They lead to distributions $R^{(i)}(\varphi)$ for the i 'th configuration in the sequences, including an arbitrary initial distribution $R^{(0)}(\varphi)$. The recursion

$$R^{(i+1)}(\varphi') = \int D\varphi R^{(i)}(\varphi) T(\varphi, \varphi') \quad (14)$$

gives in an obvious shorthand notation

$$R^{(i)} = R^{(0)} T^i. \quad (15)$$

To converge to the desired distribution (12) we clearly have to obey the stability equation

$$PT = P. \quad (16)$$

We also need ergodicity in the sense of $T^k(\varphi, \varphi') > 0$ for any pair of arguments φ, φ' and k larger than some k_0 . This implies that our algorithm can take us from any configuration to any other one in some number of steps. For most applications in field theory this is not a problem, in principle. One can always blend in some ergodic but not necessarily efficient (see below) update steps. In practice some moves called for in implementing (12) may take a forbiddingly large number of iterations. Naturally, this leads to problems to be discussed in the following sections. With stability and ergodicity and T being a probability ($T(\varphi, \varphi') \geq 0$, $\int D\varphi' T(\varphi, \varphi') = 1$) there are theorems [1] proving that beside the eigenvalue one with eigenvector P , the remaining ones are smaller than one in modulus. This guarantees convergence to P for an arbitrary start of the simulation(s).

Most of the time one enforces (16) by the sufficient but not necessary condition of detailed balance

$$P(\varphi)T(\varphi, \varphi') = P(\varphi')T(\varphi', \varphi). \quad (17)$$

A typical example is a local heatbath algorithm. One sweeps the lattice in some order with elementary transitions,

$$T = \prod_x H_x \quad (\text{some order}), \quad (18)$$

and takes for the local steps

$$H_x(\varphi, \varphi') = \text{const} \prod_{y \neq x} \delta(\varphi(y) - \varphi'(y)) P(\varphi') \quad (19)$$

with the constant determined by normalization. In words: in each local step all but one of the degrees of freedom are frozen, and the remaining one is resampled with the equilibrium distribution. Each H_x fulfills detailed balance, and this implies stability (16) for the product T .

2.3 Speed of an algorithm, critical slowing down

From the spectral properties of the update transition probabilities T one can show [1] that the distributions of configurations approach the Boltzmann factor monotonically in some suitably defined norm with a bound given by

$$\|R^{(n+1)} - P\| \leq e^{-1/\tau_{\text{exp}}} \|R^{(n)} - P\|. \quad (20)$$

The exponential autocorrelation time τ_{exp} defined here is the longest time scale in the problem. It controls after how many iterations an algorithm “forgets” the arbitrary initial configuration and approaches equilibrium. Thus, to start the estimation of observables, one has to have $n \gg \tau_{\text{exp}}$ with systematic errors being of the order $\exp(-n/\tau_{\text{exp}})$.

The question about statistical accuracy of estimation at equilibrium is a slightly different one. We have to investigate the expected discrepancy between estimate and true mean,

$$\sigma_A^2 = \left\langle \left(\frac{1}{N} \sum_{i=n+1}^{i=n+N} A(\varphi^{(i)}) - \langle A(\varphi) \rangle \right)^2 \right\rangle. \quad (21)$$

With the autocorrelation function

$$\Gamma_A(i, j) = \left\langle \left(A(\varphi^{(i)}) - \langle A \rangle \right) \left(A(\varphi^{(j)}) - \langle A \rangle \right) \right\rangle \quad (22)$$

and the equilibrium relation $\Gamma(i, j) \approx \Gamma(|i - j|)$ this is rewritten for samples much larger than τ_{exp} as

$$\sigma_A^2 = \frac{1}{N^2} \sum_{i,j=1}^N \Gamma_A(i, j) \approx \frac{1}{N} \sum_{t=-\infty}^{\infty} \Gamma(|t|) =: \frac{\Gamma_A(0)}{N} 2\tau_{\text{int},A}. \quad (23)$$

We defined the integrated autocorrelation time

$$\tau_{\text{int},A} = \frac{1}{2} + \sum_{t=1}^{\infty} \Gamma_A(t)/\Gamma_A(0). \quad (24)$$

As $\Gamma_A(0)$ is the variance of A , we see from (23) that $2\tau_{\text{int},A}$ is the number of iterations necessary to produce an effectively independent estimate for A . It is the scale relative to which the length of a certain simulation has to be judged when a calculation of A is the goal and equilibration has been achieved. If Γ is one pure exponential then $\tau_{\text{int},A}$ approximately equals that scale. In general, however, Γ has many components decaying on various time scales up to the one given by τ_{exp} . The integral quantity $\tau_{\text{int},A}$ is usually smaller than, but of a similar order of magnitude to τ_{exp} , and $\tau_{\text{int},A}$ is normally easier to determine for physical observables.

Critical slowing down amounts to the fact that typically all the autocorrelation times explode as one approaches the continuum limit of a field theoretical model. This is described by dynamical critical exponents z and z_A ,

$$\tau_{\text{exp}} \propto \xi^z \quad (25)$$

$$\tau_{\text{int},A} \propto \xi^{z_A}. \quad (26)$$

Standard local Monte Carlo algorithms have $z, z_A \approx 2$. One interprets this as being due to the growing mismatch of the physical scale ξa of (static) correlations and the lattice scale $1a$ over which the field values of neighbors influence an elementary updating move.

A natural unit for the various τ are complete passes through the lattice with a given algorithm, or sweeps. They have a computational complexity of $\text{const} \times (L/a)^D$ operations per sweep. If we estimate $\langle A \rangle$ taking the continuum limit at fixed size in physical units L , then the cost grows as

$$\text{cost} \propto (L/a)^{D+z_A}. \quad (27)$$

It should be clear now that reducing z_A for interesting quantities to one or even zero, as has become possible for some systems recently, allows us to remove the artificial cutoff much further from the physical scale and/or enhances the accuracy of our results.

3 Overrelaxation

3.1 Gaussian case

To discuss local updating in more detail we now bring out the dependence of the free field action (5) on one specific field value $\varphi(x_0)$,

$$S(\varphi) = \frac{1}{2\sigma^2} \left(\varphi(x_0) - \sigma^2 \sum_{|x-x_0|=1} \varphi(x) \right)^2 + \text{terms independent of } \varphi(x_0). \quad (28)$$

The width is given by

$$\sigma^2 = \frac{1}{m^2 + 2D}. \quad (29)$$

The local heatbath procedure in this case amounts to replacing $\varphi(x_0)$ by a new value $\varphi'(x_0)$,

$$\varphi(x_0) \rightarrow \varphi'(x_0) = \sigma^2 \sum_{|x-x_0|=1} \varphi(x) + \sigma\eta, \quad (30)$$

with a unit width Gaussian random number η . Adler proposed a generalization of heatbath for the Gaussian case [2] that had previously been known in the context of solving difference equations. He observed that

$$\varphi'(x_0) = (1 - \omega)\varphi(x_0) + \omega\sigma^2 \sum_{|x-x_0|=1} \varphi(x) + \sqrt{\omega(2 - \omega)}\sigma\eta, \quad (31)$$

leads to equilibrium for the free field for all values $0 < \omega < 2$. While $\omega = 1$ corresponds to heatbath, the range of overrelaxation (OR), $1 < \omega < 2$ will be of special interest. The limiting case $\omega = 2$ is called microcanonical. Random numbers no longer enter, and, since each local scalar field is just reflected through the average field of its neighbors, the deterministic discrete evolution conserves energy. In this case the stability property (16) is valid, but ergodicity does not hold.

Both for theoretical analysis [3] and for programming on vector or parallel machines it is advantageous to choose an order for local updates (in (18) for instance) such that a maximal number of independent ones (commuting H_x in (18)) are grouped together. This is achieved by checkerboard ordering. According to a parity function

$$(-1)^x = (-1)^{x^0+x^1+\dots+x^{D-1}} \quad (32)$$

we divide all sites into even and odd (assuming L/a even). For nearest neighbor coupled theories local updates on each of the sublattices are independent of each other. As we Fourier expand the fields,

$$\varphi(x) = L^{-D} \sum_p \tilde{\varphi}(p) e^{ip \cdot x} \quad (33)$$

the components

$$\tilde{\varphi}_{\pm}(p) = \frac{1}{2} [\tilde{\varphi}(p) \pm \tilde{\varphi}(p + (\pi, \pi, \dots, \pi))] \quad (34)$$

support fields on the respective sublattices. Momenta range over

$$p_{\mu} = \frac{2\pi}{L} n_{\mu}, \quad -\pi \leq p_{\mu} < \pi. \quad (35)$$

A full checkerboard update sweep leads to an evolution that couples only the two momentum modes in $\tilde{\varphi}_{\pm}$,

$$\tilde{\varphi}'_{\pm}(p) = (1 - \omega) \tilde{\varphi}_{\pm}(p) + \omega \lambda(p) \tilde{\varphi}_{\mp}(p) + \tilde{r}_{\pm}. \quad (36)$$

Here $\tilde{r}_{\pm}$ are the Fourier transformed random numbers and

$$\lambda(p) = \frac{2D - \hat{p}^2}{2D + m^2} \simeq 1 - \frac{m^2 + \hat{p}^2}{2D} + \dots \quad (37)$$

will be important for the rest. It depends on the lattice momenta

$$\hat{p}_{\mu} = 2 \sin(p_{\mu}/2) \doteq \frac{2}{a} \sin(ap_{\mu}/2), \quad -2 \leq \hat{p}_{\mu} < 2, \quad (38)$$

and extends over an interval such that $|\lambda(p)| \leq \lambda(0) < 1$. For

$$\chi(p) = \begin{pmatrix} \tilde{\varphi}_{+}(p) \\ \tilde{\varphi}_{-}(p) \end{pmatrix} \quad (39)$$

an update on the even sublattice followed by one on the odd, i. e. a full sweep, reads in two-component notation

$$\chi \rightarrow \chi' = M(\lambda, \omega)\chi + A(\lambda, \omega)r \quad (40)$$

with

$$M = \begin{pmatrix} 1 - \omega & \omega\lambda \\ (1 - \omega)\omega\lambda & 1 - \omega + (\omega\lambda)^2 \end{pmatrix} \quad (41)$$

and some matrix A that we will not need in detail. Upon iteration we have for the n 'th configuration

$$\chi^{(n)} = M^n \chi^{(0)} + \sum_{i=1}^n M^{n-i} A r^{(i)}, \quad (42)$$

where $r^{(i)}$ are the random numbers associated with the transition from $n - 1$ to n . For large n the first term in (42) will be seen to be negligible because of equilibration. Taking into account that successive Gaussian random numbers are independent,

$$\langle r^{(n)} r^{(m)} \rangle \propto \delta_{n,m}, \quad (43)$$

we find the decay of correlations between members of the sequence controlled by M as

$$\langle \chi^{(n+t)} \chi^{(n)} \rangle \propto |M|^t =: \exp(-t/\tau(p)). \quad (44)$$

By $|M|$ we mean the largest eigenvalue of M , which, for a converging simulation algorithm, has to come out smaller than one, of course. Component indices of χ are suppressed here. As different pairs of momenta in χ do not couple — this is of course a pathology of free fields — we have an autocorrelation time for each such pair.

It now is an easy matter to specialize to heatbath, $\omega = 1$. Eigenvalues of

$$M = \begin{pmatrix} 0 & \lambda \\ 0 & \lambda^2 \end{pmatrix}$$

are 0 and λ^2 , and this gives

$$\tau(p) = \frac{-1}{2 \ln |\lambda|} \simeq \begin{cases} \frac{D}{p^2 + m^2} & \text{for } m^2, \hat{p}^2 \ll 1 \\ O(1) & \text{for } \hat{p}^2 \not\approx 0, 4D \end{cases} \quad (45)$$

For small momenta, $p = 0$ for example, τ grows like $(ma)^{-2} = \xi^2$ (we reintroduced a that had been put to $a = 1$ during the calculation), and we see that $z = 2$. The same is true for the z_A of observables, that couple strongly enough to the low momentum modes.

Next we are ready to demonstrate how OR does better than heatbath. For general ω the eigenvalues of M are

$$\rho_{\pm} = 1 - \omega + (\omega\lambda)^2/2 \pm \sqrt{[1 - \omega + (\omega\lambda)^2/2]^2 - (1 - \omega)^2}. \quad (46)$$

It is now possible to choose ω just below 2 such that the argument under the square root becomes negative for all $\lambda(p)$, and then we have

$$|\rho_{\pm}| = |1 - \omega| \quad \text{independent of } p. \quad (47)$$

It is optimal to just start fulfilling this condition for $p = 0$ with

$$4 \frac{\omega - 1}{\omega^2} = \lambda(0)^2 \quad (48)$$

which implies close to the scaling limit $m \rightarrow 0$

$$\omega \simeq 2 \left(1 - \frac{m}{\sqrt{D}} \right) + O(m^2) \quad (49)$$

and

$$\tau(p) \simeq \frac{\sqrt{D}}{2m}. \quad (50)$$

The linear appearance of m in this formula means that we reduced z from two to one. As we just change a parameter in the local procedure, this is in practice an enormous improvement, and it comes essentially free of cost. In $D = 4$ the gain for zero-momentum observables is a factor 4ξ .

For the precise OR implementation just discussed with a tunable parameter ω not many workable extensions to non-Gaussian models have been found. A different variant, which I would like to call hybrid overrelaxation, is however applicable to many models. It is in fact at present the optimal known method for pure Yang Mills theory on the lattice. Here we only need an analogue of the microcanonical $\omega = 2$ steps. We will then update with N such steps followed by one ordinary sweep of heatbath, for instance. It will turn out that making N large in the continuum limit will produce results

similar to the tuning of ω in (49). To the author's knowledge no analysis of this variant of OR in the exactly solvable Gaussian case has been presented. This gap is presently filled [4], and some results will be shown here.

The place of $M(\lambda, \omega)$ in the iteration (42) is now taken by a matrix

$$K(\lambda, N) = M(\lambda, 1)M(\lambda, 2)^N. \quad (51)$$

The eigenvalues of K are computed by first carrying out a similarity transformation diagonalizing $M(\lambda, 2)$, whose eigenvalues are a conjugate pair of phases $\exp(\pm i\theta)$ with

$$|\lambda(p)| = \cos(\theta(p)/2). \quad (52)$$

Now the eigenvalues of K are

$$\kappa_1 = 0, \quad \kappa_2 = |\lambda| \cos((N + 1/2)\theta). \quad (53)$$

If we now keep $N\theta(0)$ fixed as $m \rightarrow 0$ such that κ_2 is close to $1/2$, for example, we effectively have $z = 1$ for the zero momentum mode. This is because the autocorrelations between configurations separated by our compound updates do not grow (for that mode), but one such iteration requires a number of sweeps that grows linearly with the correlation length. In this case we include a factor $N + 1$ in the conversion from eigenvalues to times τ , which hence always refer to single sweeps. While we focus on one mode, there are however many closely spaced values of $\theta(p)$ in a large system, and for some of them the cosine in (53) will be close to one in magnitude. That mode is then slowed down with $\tau \approx -(N + 1)/\ln |\lambda| = O(m^{-3})$. For a representative choice of parameters this is shown in Fig.1. The figure suggests that on average the τ are of order N with only isolated neighborhoods in p -space being slower. Their position depends sensitively on N . If, instead of a fixed number of microcanonical steps, we choose this number at random with average N — for instance between 1 and $2N - 1$ with equal likelihood — the picture changes. The eigenvalues of the associated matrix $\tilde{K}(\lambda, N)$ are

$$\bar{\kappa}_1 = 0, \quad \bar{\kappa}_2 = |\lambda| \cos((N + 1/2)\theta) \frac{\sin((N - 1/2)\theta)}{(2N - 1) \sin(\theta/2)}. \quad (54)$$

The corresponding plot for the same parameters as before is shown in Fig.2. A similar situation has been found [5] with hybrid Monte Carlo. In both cases the phenomenon has to do with the large number of cyclic modes in

Figure 1: Momentum dependent autocorrelation times in units of sweeps for heatbath, hybrid overrelaxation and Adler overrelaxation (eq.(48)).

free field theory. In [5] it was concluded that randomizing the microcanonical (trajectory) length remains advantageous also after inclusion of a φ^4 interaction, which was very weak, however. The author tested the randomization of N in a hybrid OR simulation of the asymptotically free CP^3 model [6] and found absolutely no difference compared to fixed N .

3.2 Overrelaxation for a compact spin system

In many interesting field theoretical models fields do not take values in linear spaces, but in compact manifolds like real spheres S^n or Lie groups like $SU(n)$. A typical family of such systems are the $O(n)$ invariant nonlinear σ -models. The partition function with the standard lattice discretization is

Figure 2: As Fig.1 for a fluctuating number of microcanonical sweeps with $N = 10$ on average.

given by

$$Z = \int \prod_x d^n s(x) \delta(s(x)^2 - 1) e^{\beta \sum_{x,\mu} s(x) \cdot s(x+\hat{\mu})}, \quad (55)$$

where $s(x) \in S^{n-1}$ has n components. For $n = 2$ this is the x-y model, which in two dimensions is expected to show a Kosterlitz Thouless transition controlled by vortices. In the same space time dimension the models with $n \geq 3$ are believed to be asymptotically free. While we will investigate the physics of the latter case in great detail in the next section, we now want to introduce hybrid OR for the $O(n)$ models.

For a local update we again focus our attention on one given site x_0 . The Boltzmann factor depends on $s(x_0)$ through

$$\exp \{s(x_0) \cdot h(x_0) + \text{terms independent of } s(x_0)\}$$

with a local magnetic field

$$h(x_0) = \beta \sum_{|x-x_0|=1} s(x). \quad (56)$$

In standard local updates this distribution of $s(x_0)$ is generated with the Metropolis or (in closed form for $n=3$) heatbath procedure. For the additional microcanonical steps in hybrid OR we may reflect the spin through a plane orthogonal to the local field,

$$s \rightarrow s' = -s + 2 \frac{h(h \cdot s)}{h^2}. \quad (57)$$

This update may be characterized as follows:

- The energy is unchanged, $h \cdot s = h \cdot s'$.
- The measure in (55) is invariant, since the reflection belongs to $O(n)$.
- Under the constraint of conserved energy, s' is as far from s as possible in a notion of distance suggested by the model, i.e. $s \cdot s'$ is minimal.

These (in the present case admittedly rather trivial) observations are useful for other cases, as one rephrases in a model independent way what we do in the $O(n)$ system.

The authors of [7] have applied hybrid OR to the $O(3)$ model in rather large scale simulations. In Fig.3 we reproduce a plot from their paper. It shows essentially the integrated autocorrelation time of the total magnetization versus N_{or}/ξ with N_{or} being the number of microcanonical steps that we called N previously. The tiny additional powers of ξ in the quantities plotted are not essential for our purpose. The fact that we see an approximately universal curve of data at widely differing correlation lengths demonstrates that z is not far from one. For instance, the cluster of points around $N_{or}\xi^{-1.1} \approx 0.8$ contains $\xi \approx 11 \dots 225$. In the paper the estimate is $z = 1.1$. The authors actually tried various combinations of microcanonical and normal updates and conclude from their data as a general strategy that a z close to the optimal one is achieved by raising N_{or} with ξ such that measurements taken in between remain with constant correlation. This coincides with our analytical findings in the free field model.

Figure 3: Plot of autocorrelation times for hybrid overrelaxation in the $O(3)$ model reproduced from ref.[7].

3.3 Pure gauge theory, $SU(2)$ and $SU(3)$

We now briefly consider lattice gauge theory where field variables $U_\mu(x) \in SU(n)$ reside on the links of our lattice. The standard Wilson action

$$S(U_\mu(x)) = \frac{\beta}{n} \sum_p \text{Re tr}(1 - U_p) \quad (58)$$

is formulated in terms of gauge invariant traces over plaquette variables

$$U_p = U_\mu(x)U_\nu(x + \hat{\mu})U_\mu(x + \hat{\nu})^{-1}U_\nu(x)^{-1}, \quad (59)$$

and the sum is over all plaquettes. It approaches the continuum Yang Mills action

$$\frac{1}{4g^2} \int d^4x F_{\mu\nu}^i(x) F_{\mu\nu}^i(x)$$

for smooth fields as the lattice spacing a goes to zero. The partition function

$$Z = \int \prod_{x\mu} dU_\mu(x) e^{-S(U)} \quad (60)$$

involves invariant integrations with the group measure on every link, and expectation values of observables are defined in the usual way. In updating a given link $x_0\mu_0$ the relevant part of the action now is

$$S(U) = -\frac{\beta}{n} \text{Re tr}(U_{\mu_0}(x_0) H_{\mu_0}^\dagger(x_0)) + \text{terms independent of } U_{\mu_0}(x_0). \quad (61)$$

In this case the environment enters through a sum over three link products (“staples”) of incomplete plaquettes after factoring out $U_{\mu_0}(x_0)$, symbolically given as

$$H^\dagger = \sum_{\text{staples}} UUU \quad (62)$$

with $2(D-1)$ terms from all directions along the axes orthogonal to the link under consideration.

The construction of microcanonical steps is now different for $n=2$ and $n=3$. The first case is easy if one observes the correspondence between $SU(2)$ and S^3 . We expand H in terms of Pauli matrices $\vec{\tau}$,

$$H = h_0 + i\vec{h} \cdot \vec{\tau}, \quad h_0^2 + \vec{h}^2 = 1 \quad (63)$$

and likewise for $U_{\mu_0}(x_0)$. Now the trace in (61) goes over to a scalar product in the component vectors like $(h_0, \vec{h})$, and the reflection (57) can immediately be generalized. Rather than spelling this out in detail we now give the translation back into $SU(2)$ language. A special feature of $SU(2)$ related to the S^3 correspondence is the fact that sums of $SU(2)$ matrices like H can always be written as

$$H = \rho V, \quad \rho \in R, \rho \geq 0, V \in SU(2). \quad (64)$$

Figure 4: Autocorrelation function for $SU(2)$ gauge theory on 32^4 with heat-bath vs. OR with $N=4$ from ref.[9].

With this polar decomposition we now have the “flip” operation [8]

$$U \rightarrow U' = VU^\dagger V, \quad (65)$$

that leaves the energy invariant, as a trivial calculation shows. The above algorithm has been tested in [9] on 32^4 lattices at two different β values. Although systematic studies of autocorrelation times are very demanding and do not exist, the autocorrelation functions of large Wilson loops in Fig.4 for both heatbath and hybrid OR with $N = 4$ very clearly show the gain that OR achieves.

For $SU(3)$ the situation is more complicated, as an analogue of (64) does not exist. A possible solution is however to still associate an $SU(3)$ element with H , namely a V , such that, when U runs through $SU(3)$,

$$\text{Re tr}(UH^\dagger)|_{U=V} = \text{maximal} \quad (66)$$

holds. With this group element constructed numerically in some approximation we then propose the substitution (65). In $SU(3)$ this in general results in an energy change, and a nontrivial accept/reject step has to follow as with Metropolis updating. This procedure coincides with the previous one for $SU(2)$. If the “sum over staples” is nearly proportional to a group element, as one naively expects for large β , the acceptance should be very high and the update almost microcanonical. Note that an error in solving (66) does not lead to systematic errors in the simulation but just influences the acceptance. In ref.[10] an acceptance of more than 99% was indeed found for physically interesting β in $SU(3)$ lattice gauge theory. Compared with a number of algorithms including pseudo-heatbath, Metropolis and hybrid Monte Carlo, OR was found to be most efficient. Again, however, the study was not systematic but just compared a few quantities on even rather small lattices. Also the implementation was somewhat different here: one almost microcanonical step of the type just described was followed *locally* with one (presumably not too large) Metropolis step. (Note that many hits or any other good approximation to heatbath would actually “undo” the OR step.) A different possibility from the one described here would be microcanonical updates in $SU(2)$ or even $U(1)$ subgroups that one embeds locally into the $SU(3)$ links.

To summarize, we think that OR is “state of the art” at present for simulations of 4-D pure gauge theory. The recently emerged tuning strategy of choosing N proportional to the scale should be employed. With OR and rapidly developing parallel computing facilities, for which OR is ideally suited, we may soon see results which are not so far away from what can be done in two dimensions. A first result of this kind is an impressive plot of the interquark force in $SU(2)$ from ref. [11] reproduced in Fig.5. The calculation is on a $48^3 56$ lattice, and the scale associated with the string tension in lattice units is about 15.

Figure 5: Interquark force for $SU(2)$ gauge theory from ref.[11].

4 Cluster algorithm and $O(3)$ σ -model results

4.1 Ising model, Swendsen Wang algorithm

Cluster algorithms were first developed[12] for the Ising model

$$Z = \sum_{\{\sigma(x)=\pm 1\}} e^{\beta \sum_{x\mu} \sigma(x)\sigma(x+\hat{\mu})}. \quad (67)$$

As this algorithm appears as one component in the updates for σ -models that we want to study here, we introduce it first. Due to the simple nature of the Ising field variables, the only possible moves are sign-flips. Therefore

a general nonlocal algorithm has to construct domains C , such that

$$\sigma(x) \rightarrow -\sigma(x) \big|_{\text{all } x \in C} \quad (68)$$

are appropriate collective moves for (67). Swendsen and Wang achieve this with the help of a percolation process. We derive it by introducing additional variables into (67) without changing it, apart from overall factors that we do not keep track of for simplicity (they cancel in expectation values).

As a first step note that

$$Z = \sum_{\{\sigma(x)\}} \prod_{x\mu} \left[(1-p) + p\delta_{\sigma(x),\sigma(x+\hat{\mu})} \right] \quad (69)$$

has the same relative weight for configurations as in (67), provided we put

$$p = 1 - e^{-2\beta}. \quad (70)$$

We then introduce a field $k_\mu(x) = 0, 1$ to enumerate the two terms on every link

$$Z = \sum_{\{\sigma(x), k_\mu(x)\}} \left(\frac{p}{1-p} \right)^{\sum_{x\mu} k_\mu(x)} \Delta(k|\sigma), \quad (71)$$

where all the δ symbols from (69) are combined in

$$\Delta(k|\sigma) = \begin{cases} 0 & \text{if on any link } \sigma(x)\sigma(x+\hat{\mu})k_\mu(x) = -1 \\ 1 & \text{else} \end{cases}. \quad (72)$$

It enforces a constraint that bonds can only be “active” ($k_\mu(x) = 1$) between parallel spins and vice versa that spins connected by such bonds have to be aligned. The algorithm now consists of alternating updates of spins and of bonds. With given spins, the bonds can be chosen independently of their previous values to be

$$k_\mu(x) = \begin{cases} 1 & \text{with probability } p\delta_{\sigma(x),\sigma(x+\hat{\mu})} \\ 0 & \text{else} \end{cases}. \quad (73)$$

This corresponds to (perfect) independent sampling for this subset of the variables. For given bonds, one decomposes the sites into a maximal number of clusters such that no pair of sites in two different clusters is connected by

an active bond. This step also appears in the study of bond percolation [13], and it can be carried out with a number of operations proportional to the volume. Then (71) instructs us to assign to each cluster as a whole (i.e. to all its spins in parallel) one of the two possible spin orientations with equal probability. In this step a nonlocal rearrangement of the spins takes place, to which the good efficiency ($z \approx 0.3$), that has been found numerically for Swendsen Wang updating in the Ising model, has to be attributed.

One can moreover prove an identity [14], that exhibits the connection between the clusters and spin physics,

$$\langle \sigma(x)\sigma(y) \rangle = \left\langle \sum_{\text{clusters } c} \Theta_c(x)\Theta_c(y) \right\rangle \quad (74)$$

with $\Theta_c(x)$ the characteristic function of a cluster. The relation demonstrates that in a well defined sense the clusters have a size coupled to the correlation length. In particular, they percolate in the magnetized phase of the Ising model, and the threshold coincides with the critical point. The relation also has practical relevance, since it allows us to estimate correlations with a reduced variance.

A closely related technique has been proposed in [14] in the form of the single cluster algorithm. There one picks a random site x_0 (the “seed”) and grows a cluster out of that site. Only those bonds are examined that one meets in this growth process, and they are activated as before. When the growth stops, all spins on the one cluster formed are flipped. This easily programmable version has been found more efficient, in particular in more than two dimensions. In $D = 4$ it seems to give $z = 0$ whereas the original many cluster technique has $z = 1$ [15]. Results in three dimensions are $z_E = 0.28$ and $z_\chi = 0.14$ for energy and squared total magnetization respectively [16].

4.2 Embedded Ising variables

Sometime after the discovery of the cluster algorithm for Ising and Potts models by Swendsen and Wang it was realized that it can also be applied to systems with continuous spin manifolds through embedding. A simple example is φ^4 with action (5) augmented by a quartic local interaction term. There one may introduce the sign of the field as an explicit degree of freedom

[17]. Starting from a given configuration of the simulation one writes

$$\varphi(x) = |\varphi(x)|\sigma(x). \quad (75)$$

The cluster update of $\{\sigma(x)\}$ proceeds with an effective action

$$S_{\text{eff}}(\sigma) = - \sum_{x\mu} |\varphi(x)| |\varphi(x+\hat{\mu})| \sigma(x) \sigma(x+\hat{\mu}) + \text{terms independent of } \sigma. \quad (76)$$

This represents a slightly generalized Ising system (at fixed $|\varphi|$) with bond dependent ferromagnetic couplings. Cluster algorithms can be applied to it by simply substituting these coupling values into (70). The cluster update for the sign of φ has to be supplemented with other steps that can also change the magnitudes to make the compound ergodic. It is plausible, however, that at large quartic coupling, where φ^4 in the critical region has a double well structure, the behavior is Ising-like and the sign fluctuations are most important. In this region it is indeed found in ref.[17] that the cluster steps significantly accelerate the simulation.

Also the $O(n)$ nonlinear σ -model (55) may be simulated with Ising embedding [14]. In this case we first pick a random spin direction $r \in R^n$, $|r| = 1$, independently of the spin field $s(x)$ and decompose

$$s(x) = \sigma(x)|r \cdot s(x)|r + s_{\perp}(x). \quad (77)$$

Ising flips are now mapped to reflections of $O(n)$ spins through the plane orthogonal to r . As any two n -component unit vectors can be connected by a reflection through some plane, these updates alone are even ergodic, if all directions can appear for r . It is clearly natural to choose r with the same uniform measure on the sphere as the spins themselves. The effective embedded Ising action has again ferromagnetic random bond strengths,

$$S_{\text{eff}}(\sigma) = - \sum_{x\mu} |r \cdot s(x)| |r \cdot s(x+\hat{\mu})| \sigma(x) \sigma(x+\hat{\mu}) + \text{terms independent of } \sigma. \quad (78)$$

All applications (both single and many cluster) of this algorithm have yielded results which are consistent with the complete absence of critical slowing down, $z = 0$ [18]. Typical bounds quoted are $z \leq 0.1$ [19]. The autocorrelation times for typical observables, in particular with single cluster updating, just stay constant and are typically even smaller than one. This means that

building and reflecting a number of clusters such that the computational effort corresponds to one pass through the lattice, achieves complete decorrelation. Moreover, variance reduction — as discussed in the Ising case — is also available for the σ -models. In this very favorable situation it is possible to get accurate results close to the continuum limit, with relatively modest computer time, as will be described in the following.

4.3 S-matrix of $O(n)$ σ -models

For the $O(n)$ σ -models an exact S-matrix has been proposed by Zamolodchikov in ref.[20]. Before we turn to numerical methods for computing scattering phases, we now briefly sketch the arguments that lead to the exact S-matrix. One starts with the standard scenario for the two dimensional model: asymptotic freedom and mass generation with an $O(n)$ vector multiplet of massive fundamental particles. They are associated with one particle states $|p, a\rangle$, to which the fundamental spin $s(x)$ field couples,

$$\langle 0|s^a(x)|p, b\rangle = \delta^{ab}Z^{1/2}e^{ip\cdot x}, \quad (79)$$

and a, b are $O(n)$ indices here, and p is the (on shell) momentum. This standard scenario is supported by the perturbative renormalization group and by the $1/n$ expansion. It has also been criticized in a series of papers by Patrascioiu and Seiler, see for instance [21]. It is clearly of interest to see how far the standard picture is consistent with numerical results, although a real proof will never result from this.

For the S-matrix it is of particular importance that beyond the features just mentioned, the $O(n)$ models possess an infinite number of conserved (nonlocal) charges. This leads to the absence of particle production, and the set of outgoing momenta is restricted to be identical to the incoming ones. Moreover, there are identities that allow to express arbitrary matrix elements in terms of 2-particle scattering. In [20] this was combined with general principles like unitarity, crossing and analyticity, and it turned out that then the S-matrix for a massive n -plet is almost fixed. The simplest of the remaining choices agrees with perturbation and $1/n$ expansion and is proposed as the exact S-matrix.

The Zamolodchikov 2-particle S-matrix can be decomposed into isospin channels $I = 0$ (scalar), $I = 1$ (antisymmetric tensor) and $I = 2$ (symmetric

Figure 6: Exact scattering phases in the 2-particle isospin channels of the $O(3)$ model.

tensor) [22]. For each of them there is a real phase δ_I . Their dependence on the momentum transfer is conveniently given in terms of the rapidity difference $\theta = \eta - \chi$, where the momentum p of one incoming particle is

$$p = \begin{pmatrix} m \cosh \eta \\ m \sinh \eta \end{pmatrix}, \quad (80)$$

and there is a corresponding expression for the other momentum in terms of χ . The phases for the case $n = 3$ are given by

$$e^{2i\delta_0} = \frac{\theta + 2\pi i}{\theta - 2\pi i} \quad (81)$$

$$e^{2i\delta_2} = \frac{\theta - \pi i}{\theta + \pi i} \quad (82)$$

$$e^{2i\delta_1} = e^{2i(\delta_0 + \delta_2)}. \quad (83)$$

This momentum dependence of δ_I is shown in Fig.6. All phases vanish logarithmically for large momenta in accordance with asymptotic freedom.

There is no known practical way to simulate real scattering in Minkowski space. Lüscher noticed however [23] interesting quantitative relations between the S-matrix and leading finite volume effects in euclidean space for $mL < \infty$. This is intuitively plausible in the naive sense that wave packets associated with outgoing particles in scattering states in infinite volume have some probability to overlap again and again on a torus. For the case at hand, Lüscher's method requires information about the spectrum of 2-particle states (the lowest states even under charge conjugation $s^a(x) \rightarrow -s^a(x)$) in the various isospin channels at zero total momentum. Where there would be a continuum in infinite volume, there are densely spaced discrete levels for finite L . The extension T in euclidean time should be effectively infinite compared to L to have zero temperature and to be able to resolve the spectrum of interest by analyzing time decay properties. For each such energy W found, one has to compute k by putting

$$W = 2\sqrt{m^2 + k^2}. \quad (84)$$

Then a point of the graph $(k, \delta_I(k))$ is given by

$$e^{2i\delta_I(k)} = e^{-ikL}. \quad (85)$$

In free theory with vanishing phases (85) would just be the usual periodicity condition for the equal and opposite momenta of the two particles. For the modification in the interacting case to hold, it is important that polarization effects due to selfinteractions of the particles around the torus have to be negligible. This can be arranged for short range interactions. For the σ -model these interactions are known to be exponentially small in mL [24]. Formula (85) can be made particularly transparent in the nonrelativistic limit. We refer the reader to [22] for such a discussion.

We now come to numerical results using the finite size method for scattering phase shifts. A typical run with the single cluster method in ref.[22] was at parameters $L = 128$, $T = 256$, $\xi = (ma)^{-1} = 13.632(6)$. For a first orientation we checked the influence of cutoff effects on the energy momentum dispersion relation in the 1-particle sector of the spectrum. For spatial momentum $p^1 = 2\pi\nu/L$ we determined the corresponding energy $\hat{p}^0$ from decay in euclidean time and compared with $p^0 = \sqrt{m^2 + (p^1)^2}$. Some results are in the following table:

ν	p^1/m	$\hat{p}^0/p^0$
1	0.67	0.9992(3)
2	1.34	0.9993(3)
$\vdots$	$\vdots$	$\vdots$
8	5.35	0.9904(4)

The discrepancy visible in the last column can be fitted as $1 - 0.06(p^1 a)^2$, a typical small cutoff effect of the type that is expected in physical observables [25]. It grows with momentum, as waves with larger wave numbers feel the discretization more and more. But we see that for a rather relativistic momentum at $\nu = 8$ we still only have effects at the percent level.

To access levels in the 2-particle sectors we measured the matrix of correlations

$$C_{pp'}^I(t) = \langle O_I^{ab}(0, p) O_I^{ab}(t, p') \rangle, \quad (86)$$

with composite fields

$$O_I^{ab}(t, p) = P_I(ab|cd) \sum_{xy} e^{ip(x-y)} s^c(t, x) s^d(t, y). \quad (87)$$

In these formulas p, p', x, y are purely spatial, a, b, c, d are $O(n)$ indices, and we use isospin projectors P_I , which for $I = 2$ for instance, reads

$$P_2(ab|cd) = \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}).$$

A somewhat involved procedure, for whose details we refer to [22], now allows us to derive a number of levels from the correlation matrix C . They are converted to phases by (84) and (85) and shown together with the exact curves (81),(82),(83) in Fig.7. The various symbols refer to simulations at different parameters. The filled symbols stem from the run mentioned before with $L = 128$ and $(ma)^{-1} \approx 13.6$. The open symbols refer to an approximately two-fold larger lattice spacing at the same physical size mL . As expected they have a tendency to deviate more from the continuum curves at larger momenta. For the simple error bars we used a physically smaller lattice to verify that our results are insensitive to the polarization effects mentioned before.

Figure 7: Numerically determined scattering phases in the 2-particle isospin channels $I = 0, 1, 2$ of the $O(3)$ model together with curves predicted by the Zamolodchikov S-matrix (see text for an explanation of the symbols).

We see a close coincidence between theoretical prediction and numerical results with small discrepancies characteristic of finite lattice spacing. In principle, with more data than those shown here, one could attempt a systematic extrapolation to the continuum for this model. We feel, however, that the existing data already leave little doubt that the method works, and that the Zamolodchikov S-matrix indeed describes the $O(3)$ model. The latter is not entirely trivial, as the derivation, as we mentioned, involves various assumptions including one mainly based on simplicity. The numerical calculation required rather precise data on 4-point functions, which could be obtained with reasonable CPU-effort by the cluster technique.

4.4 Some properties of asymptotically free field theories

In this section we summarize some properties of asymptotically free theories like $O(n)$ invariant spin models in two and pure gauge theories in four dimensions. Although well known to particle physicists, they are decisive for a discussion of scaling and running coupling “constants” that follows. All the models — whether defined formally in the continuum or on a lattice — have at first sight only one free parameter, the bare coupling g_0^2 . On the lattice it is often given by the analogous inverse temperature, $\beta = 1/g_0^2$ for the spin models or $\beta = 2n/g_0^2$ for $SU(n)$ gauge theory. As an implicit scale the cutoff has to be added, and we will restrict ourselves here to the lattice with spacing a . In the regularized theory one computes physical scales that we generically call m_i , and they appear as dimensionless ratios

$$m_i a = f_i(g_0^2) = 1/\xi_i. \quad (88)$$

The continuum limit is reached at a critical coupling, $g_0 \rightarrow g_{0c}$, where all correlation lengths ξ_i diverge. We interpret this as the spacing a vanishing with some mass, say m_0 , staying finite. This is then the only parameter in the continuum theory. This kind of limit is called dimensional transmutation in particle physics. The critical bare coupling is $g_{0c} = 0$ according to the standard picture based on perturbation theory. It should be clear that both the lattice scale a and the distinguished mass m_0 are only (convenient) bookkeeping devices. In principle one could restrict oneself to just talking about dimensionless universal combinations like

$$\frac{m_i}{m_j} \Big|_{\text{continuum}} = \lim_{g_0 \rightarrow 0} \frac{f_i}{f_j} = \lim_{g_0 \rightarrow 0} \frac{\xi_j}{\xi_i}. \quad (89)$$

Apart from direct physical scales like masses, there is a class of quantities in asymptotically free models that have an interpretation as renormalized coupling constants. A typical example is some scattering cross section depending on a euclidean momentum transfer Q^2 . After removing kinematical factors such that the perturbation expansion starts with the bare coupling, one has in the regularized (lattice) theory

$$g^2(Q) = g_0^2 + \sum_{k=2}^{\infty} c_k(Qa) g_0^{2k}. \quad (90)$$

We can also employ such a quantity to fix and parameterize the continuum limit by keeping $g(Q_0)$ at a certain numerical value and sending $a \rightarrow 0$ by tuning g_0 critical. Now Q_0 takes the rôle of a continuum scale in place of a mass like m_0 . It should be emphasized that many such quantities are possible. One may use various cross sections, and, as we shall see later, there are also very different scale dependent (“running”) couplings having to do with the volume dependence on the torus. What they all have in common is the existence of a formal expansion like (90) and their dependence on a physical scale staying finite (relative to all other physical scales) in the cutoff limit. If we have two such couplings $g^{(1)}(Q)$ and $g^{(2)}(Q)$ one may reorganize (90) to an expansion yielding in the cutoff limit

$$\left(g^{(1)}(Q)\right)^2 = \left(g^{(2)}(\tilde{Q})\right)^2 + \sum_{k=2}^{\infty} d_k^{(1,2)}(Q/\tilde{Q}) \left(g^{(2)}(\tilde{Q})\right)^{2k}. \quad (91)$$

Renormalizability guarantees all the coefficients $d_k^{(1,2)}$ to exist and remain finite in the continuum limit. Asymptotic freedom implies that when $Q, \tilde{Q} \gg m_0$ — at physically high energy or short distance — the renormalized perturbation expansion (91) can be meaningful (for not too large $|\ln(Q/\tilde{Q})|$) since both couplings are small. Therefore it is in principle sufficient to know one running coupling at high energy, because one can then relate it to others analytically by perturbation theory. There is thus a whole perturbative sector of the theory. Its connection to nonperturbative scales like the massgap m_0 is a typical problem for numerical calculations. There one could, for instance, seek the answer to a question like: at which ratio Q/m_0 does a certain coupling $g^2(Q)$ reach a certain value like 0.1? With one such information the scale of the whole perturbative sector is set relative to a nonperturbative scale.

Among the perturbatively computable quantities is the change of g with Q at high energy,

$$Q \frac{dg^2}{dQ} = \beta(g^2). \quad (92)$$

It can be considered as an infinitesimal version of (91) with both couplings identical and infinitesimally differing Q and $\tilde{Q}$. The perturbative renormalization group β -function is given as

$$\beta(u) = -u^2 \sum_{l=0}^{\infty} b_l u^l \quad (93)$$

with b_0 and b_1 universal for differently defined couplings. These two coefficients are usually known, while the first nonuniversal one b_2 is only known for a few models and couplings. The negative sign of β (asymptotic freedom) for small arguments implies that any renormalized coupling, once perturbative evolution is applicable at high energy, vanishes asymptotically as

$$g^2(Q) \simeq \frac{b_0}{\ln(Q)} \quad \text{as } Q \rightarrow \infty. \quad (94)$$

Defining the continuum limit by keeping some renormalized coupling fixed also determines the rate at which the bare coupling vanishes with the lattice spacing. It is given by the bare β -function $\bar{\beta}$ in

$$a \frac{dg_0^2}{da} = -\bar{\beta}(g_0^2), \quad (95)$$

and it again has an expansion like in (93) with the universal b_0, b_1 and some $\bar{b}_2, \dots$. This simply follows from the constancy of the renormalized coupling combined with (90), (92) and simple dimensional analysis. The differential equation (95) fixes all functions f_i in (88) up to a multiplicative constant which is *beyond* perturbation theory,

$$f_i(g_0^2) \propto (b_0 g_0^2)^{-b_1/b_0^2} \exp(-1/b_0 g_0^2) (1 + O(g_0^2)). \quad (96)$$

From this asymptotic behavior of all masses in lattice units one can define an artificial perturbative scale Λ_L which in the two loop approximation is given by

$$\Lambda_L^{(2)} = \frac{1}{a} (b_0 g_0^2)^{-b_1/b_0^2} \exp(-1/b_0 g_0^2). \quad (97)$$

Masses can now be referred to this scale by

$$\frac{m}{\Lambda_L} \Big|_{\text{continuum}} = \lim_{g_0^2 \rightarrow 0} \frac{m(g_0^2)}{\Lambda_L^{(2)}} \quad (98)$$

One speaks of asymptotic scaling when the right hand side looks constant to some accuracy and can hence be extrapolated to $g_0^2 = 0$.

Asymptotic scaling is the traditional way to try to establish the connection between nonperturbative and perturbative continuum physics. The convergence rate is proportional to g_0^2 or a higher power of it, if Λ_L can be

Figure 8: Ratio of massgap to perturbative Λ_L -scale in the two and three loop approximation. Asymptotic scaling would correspond to a constant ratio. The data shown range over correlation lengths from $\xi = 6.9$ to $\xi = 122$.

approximated by more loops. Because of the exponential growth of correlation lengths with vanishing g_0 implied by (96) it is hardly possible to vary $g_0^2 \propto 1/\ln \xi$ by significant amounts. This led to asymptotic scaling being an essentially untested hypothesis in many older studies with large errors. The new techniques for two dimensional spin models allow now some variation of $\ln \xi$ with much more accurate results. It turned out that, rather than allowing a reliable extrapolation $g_0^2 \rightarrow 0$, the variation is just enough to show that this is numerically hopeless in those models. A demonstration for the $O(3)$ model [26] is given in Fig.8. There we also see as a horizontal line what is proposed to be the exact value in [27]. No smooth asymptotic approach to a constant value is visible, and any extrapolation of the data is clearly impossible.

The situation in the $SU(3) \times SU(3)$ and CP^3 spin models, that could recently be studied in detail due to progress with multigrid methods [40], is just the same as $O(3)$. To conclude, let us emphasize the all important difference between asymptotic scaling and scaling. The latter means convergence proportional to a^2 or, in less favorable cases, a (inverse powers of ξ) rather than logarithmic. An example was the relativistic dispersion relation and the scattering phases presented earlier.

4.5 Numerical estimation of physical running couplings

We now outline a method to compute physical running coupling constants at high energy including the normalization of energy relative to the nonperturbative sector. It has been carried out for the $O(3)$ model [28] and relies only on the verification of scaling with convergence proportional to powers of the lattice spacing. The development of techniques for such a computation is attractive to ultimately compare with experimentally measured values of the strong QCD coupling constant $\alpha_s(M_Z) \approx 0.11$. Here the mass M_Z of the Z-particle is, from the point of view of QCD, just some high energy scale, about 200 times that given by the string tension or 118 times the ρ -mass. From what was presented in the last section it is clear that such a number — in principle — is calculable without free parameters in asymptotically free theories.

As different physical couplings can be connected by perturbation theory at high energy, we choose a coupling for our computation guided by technical convenience. The decisive point is to take the finite spatial size L as the (only) free scale parameter, on which the coupling depends. Then universal continuum results only require $a \ll L$ to hold, and one does not have to *simultaneously* arrange $a \ll Q^{-1} \ll L$ for another scale. The finite size coupling used for the $O(3)$ model is

$$\bar{g}^2(L) = m(L)L, \quad (99)$$

where $m(L)$ is the finite volume massgap [29]. It is defined from the temporal decay of spatial 0-momentum Green functions as before. While $\bar{g}^2$ is indeed a perturbative coupling with an expansion (90) set up for small systems [29],

$m(\infty)L \ll 1$, it is well defined for all L and diverges like

$$\bar{g}^2 \simeq m(\infty)L \quad \text{as } L \rightarrow \infty \quad (100)$$

in the nonperturbative regime.

With a convenient coupling chosen, the next important point is to break the problem of connecting low and high energy into several steps. This is done by introducing an analogue of β -functions for finite scale factors, the step scaling function

$$\sigma(s, u) = \bar{g}^2(sL)|_u = \bar{g}^2(L). \quad (101)$$

Its relation to the β -function is

$$\ln s = - \int_u^{\sigma(s,u)} \frac{dv}{\beta(v)}. \quad (102)$$

We emphasize here that $\sigma(s, u)$ is a concept on the renormalized level independent of the lattice. Let us imagine we knew the exact step scaling function, say for scale factor $s = 2$. We could then start at a scale which we expect to be perturbative, for instance L_1 where $\bar{g}^2(L_1) = g_1^2 = 0.9176$, to use numerical examples from [28]. Now, using σ , we iterate in the following way:

$$\begin{aligned} \sigma(2, g_1^2) &= \bar{g}^2(2L_1) &= g_2^2 > g_1^2 \\ \sigma(2, g_2^2) &= \bar{g}^2(4L_1) &= g_3^2 > g_2^2 \\ \vdots & & \vdots \\ \sigma(2, g_k^2) &= \bar{g}^2(2^k L_1) &= g_{k+1}^2 > g_k^2 \\ &= m(2^k L_1) 2^k L_1 \\ &\approx m(\infty) 2^k L_1 \end{aligned}$$

Here the first few steps can be used to check if we really started in the perturbative range. If true, this evolution should agree with a perturbative approximation of the step scaling function in (102). After some number of steps one arrives at long distance (bottom of the iteration scheme). Here we can make contact with a nonperturbative scale. In the present example this

is particularly easy as it appears in the coupling $\bar{g}^2$ itself. We finally learn what L_1 we started with in units of the inverse massgap, namely

$$L_1 = g_{k+1}^2 2^{-k} m(\infty)^{-1}. \quad (103)$$

At this point the goal of knowing a running coupling at high energy with an absolutely normalized scale relative to a mass has been achieved. It remains now to describe how we are able to compute the step scaling function nonperturbatively.

At finite lattice spacing there is an approximate step scaling function Σ with

$$\sigma(s, u) = \lim_{a \rightarrow 0} \Sigma(s, u, a/L). \quad (104)$$

A sequence of $\Sigma(2, u, a/L)$ is now constructed with fixed u and shrinking a/L . We start by tuning the bare coupling g_0 on some lattice of, for instance, $L = 10a$, such that $\bar{g}^2 = u$ is met. Then with the *same* g_0 , a simulation on $L = 20a$ gives a renormalized charge $\Sigma(2, u, 1/10)$. If this is repeated for many pairs of lattices, for example $L/a = (7, 14), (8, 16), (10, 20), \dots$, then the resulting values for Σ can be extrapolated to the continuum. The convergence rate expected for the coupling (99) is $(a/L)^2$ (neglecting the variation of logarithms). This is shown in an example from [28] in Fig.9. In Fig.10 five data points constructed by the above procedure are shown together with the perturbative evolution starting from the point at shortest distance. We see that the evolution rapidly breaks away from the perturbative curves only for scales L approaching $m(\infty)^{-1}$.

While each value of a running coupling like $\bar{g}^2(L)$ implicitly defines a scale, one can also associate a finite Λ -parameter scale with the known short distance asymptotics. Similarly to the bare charge on the lattice we define

$$\Lambda^{(2)}(L) = \frac{1}{L} (b_0 \bar{g}^2)^{-b_1/b_0^2} \exp(-1/b_0 \bar{g}^2) \quad (105)$$

and

$$\Lambda = \lim_{L \rightarrow 0} \Lambda^{(2)}(L). \quad (106)$$

As is well known, Λ parameters belonging to different couplings and also those of minimal subtraction and of the bare lattice charge can be exactly related by 1-loop calculations. This is plausible, as they are obtained from

Figure 9: Extrapolation of the step scaling function to the continuum.

the behavior of infinitesimal couplings at infinite energy. Now the convergence rate from higher loop corrections in (105) is proportional to powers of $\bar{g}^2(L)$. The situation for the scale connected with our charge (99) is shown in Fig.11, and the dashed line is again from [27]. Also here a reliable extrapolation to $\bar{g}^2 = 0$ is not possible. On the other hand it is clear from Fig.10 that the coupling at $L = 100m^{-1}$ has been very well determined. This shows, that Λ -parameters, both for bare and renormalized couplings, are not very convenient parametrizations for numerical work. It seems better, with methods like the one presented, to avoid them and to compute directly quantities more closely related to experimental numbers like $\alpha_s(M_Z)$. In [28] also the step of converting from $\bar{g}^2$ to the coupling defined by minimal subtraction (independent of a finite volume) has been carried out for the $O(3)$ model. The result is

$$\bar{g}_{MS}^2(\mu = 1/L) = \bar{g}^2(L) + 0.156 \bar{g}^4(L) - 0.035 \bar{g}^6(L), \quad (107)$$

which looks reasonably stable for the smallest coupling $\bar{g}^2 = 0.74$ that was used in our rescaling steps.

At present a similar method is developed for four dimensional pure gauge

Figure 10: Running of the coupling constant $\bar{g}^2(L)$ with the scale in units of the infinite volume massgap. The data points have been obtained numerically by extrapolation to the continuum.

theory [30]. Since small volume massgaps are not so easily measurable in this case, a different finite volume physical quantity has been proposed there. It is an adaptation of background field methods to finite volume [31]. The physical method behind it is essentially an estimate of the change in free energy from varying a dimensionless parameter associated with Dirichlet boundary conditions in one direction. First numerical tests in $SU(2)$ look positive and will hopefully be reported on in the near future.

5 Outlook

Considerable effort has been devoted to the generalization of cluster algorithms beyond ferromagnetic Ising and Potts models and $O(n)$ invariant spin systems. Some nice successes have been scored for models of interest to condensed matter physics like SOS [32], interface roughening [33], fully frustrated Ising [34] and antiferromagnetic Potts [35] models. Interesting cluster sim-

Figure 11: Ratio of massgap to the Λ -scale associated with a renormalized coupling in the two and three loop approximations.

ulations of gauge invariant models have so far been restricted to those with Ising variables on links. An elegant method [36] was proposed and tested in $D = 3$, and another variant [37], although less efficient for three dimensions, has the virtue of being generalizable to $D = 4$ and to the inclusion of Higgs fields [38]. Embedded subsystems in $U(1)$ [39] or $SU(n)$ gauge theories, for which the $Z(2)$ algorithms could be used, have not been found so far.

Also the more modest goal of extending cluster techniques to other non-linear σ -models like CP^n or chiral $SU(n) \times SU(n)$ systems has not been met. Just recently, however, Hasenbusch, Mack and Meyer have reported on a multigrid algorithm [40], that was shown to have dynamical exponents $z \approx 0.2$ at least effectively for correlation lengths up to about 100 in $O(3)$, CP^3 and $SU(3) \times SU(3)$ spin models. They propose block moves of a smooth, for instance pyramidal, shape in a hierarchical fashion on all scales. The

smoothness, as opposed to step function like moves, is absolutely crucial to get acceptance for physically relevant updates. The implementation of their multigrid updates proceeds in the unigrid fashion, where one goes back to the original fine lattice from every level of updates to compute the change in energy. This has the slight disadvantage of costing an extra $\log(L/a)$ factor in the operations count, but it is relatively easy to implement. Thus, multigrid techniques seem to be a more versatile tool than cluster algorithms, and a lot of research is devoted to their generalizations to gauge theory. One should bear in mind, however, that the crossover in efficiency between multigrid and the simple hybrid overrelaxation algorithms in spin systems is only reached at correlation length 20 or so. Although this is clearly model as well as program dependent, OR may be the method of choice for pure gauge theory for quite a while.

Finally we would like to mention a different area, where multigrid is developed vigorously, namely for the inversion of the lattice Dirac operator in a given gauge field. This problem consumes the bulk of computing power in computations involving quark fields, and there is slowing down, too. Again, there are optimized more conventional methods like conjugate gradient, which should eventually, for very large scale applications on parallel machines, be outperformed by nonlocal methods like multigrid.

Acknowledgements

I would like to thank my co-workers in the presented and ongoing calculations, Martin Lüscher, Rajamani Narayanan, Rainer Sommer and Peter Weisz, for the enjoyable and exciting collaboration.

References

- [1] A. D. Sokal, *Monte Carlo Methods in Statistical Physics: Foundations and New Algorithms*, Lecture Notes of the Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne 1989
- [2] S. Adler, Phys. Rev. D23 (1981) 2901
- [3] H. Neuberger, Phys. Rev. Lett. 59 (1987) 1877
- [4] U. Wolff, *Dynamics of hybrid overrelaxation in the Gaussian model*, preprint CERN-TH 6408/92
- [5] P. Mackenzie, Phys. Lett. 226B (1989) 369
- [6] U. Wolff, *Scaling topological charge in the CP^3 spin model*, preprint CERN-TH 6407/92
- [7] J. Apostolakis, G. C. Fox and C. F. Baillie, Nucl.Phys. B (Proc. Suppl.) 20 (1991) 678
- [8] M. Creutz, Phys. Rev. D36 (1987) 515;
F. R. Brown and T. J. Woch, Phys. Rev. Lett. 58 (1987) 2394
- [9] K. Decker and Ph. de Forcrand, Nucl.Phys. B (Proc. Suppl.) 17 (1990) 567
- [10] R. Gupta, G. W Kilcup, A. Patel, S. R. Sharpe and Ph. de Forcrand, Mod. Phys. Lett. A3 (1988) 1367
- [11] S. P. Booth et al. (UKQCD Collaboration), *Towards the continuum limit of $SU(2)$ lattice gauge theory*, Liverpool preprint LTH 271 (1991)
- [12] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. 58 (1987) 86
- [13] For a review, see, e.g., D. Stauffer, Phys. Rep. 54 (1979) 1
- [14] U. Wolff, Phys. Rev. Lett. 62 (1989) 361
- [15] P. Tamayo, R. C. Brower and P. W. Klein, J. Stat. Phys. 58 (1990) 1083
- [16] U. Wolff, Phys. Lett. 228B (1989) 379.

- [17] R. C. Brower and P. Tamayo, Phys. Rev. Lett. 62 (1989) 1087
- [18] Ch. Frick, K. Jansen and P. Seufferling, Phys. Rev. Lett. 63 (1989) 2613
- [19] R. Edwards and A. Sokal, Phys. Rev. D40 (1989) 1374
- [20] A. B. Zamolodchikov and Al. B. Zamolodchikov,
Nucl. Phys. B133 (1978) 525
- [21] A. Patrascioiu and E. Seiler, *The Difference between Abelian and Non-Abelian models: Fact and fancy*, München preprint MPI-PH-91-87 (1991)
- [22] M. Lüscher and U. Wolff, Nucl. Phys. B339 (1990) 222
- [23] M. Lüscher, Commun. Math. Phys. 105 (1986) 153
- [24] M. Lüscher, *On a relation between finite size effects and elastic scattering processes*, lecture given at Cargèse (1983), in *Progress in gauge field theory*, eds. G. 't Hooft et al. (Plenum, New York, 1984);
Commun. Math. Phys. 104 (1986) 77
- [25] K. Symanzik, Nucl. Phys. B226 (1983) 187, 205;
and *Cutoff dependence in lattice ϕ_4^4 theory* in *Recent developments in gauge theories* (Cargèse 1979) eds. G. 't Hooft et al. (Plenum, New York, 1980)
- [26] U. Wolff, Nucl. Phys. B334 (1990) 581; Phys. Lett. 222B (1989) 473
- [27] P. Hasenfratz, M. Maggiore and F. Niedermayer,
Phys. Lett. 245B (1990) 522;
P. Hasenfratz and F. Niedermayer, Phys. Lett. 245B (1990) 529
- [28] M. Lüscher, P. Weisz and U. Wolff, Nucl. Phys. B359 (1991) 221
- [29] M. Lüscher, Phys. Lett. 118B (1982) 391
- [30] M. Lüscher, R. Narayanan, P. Weisz and U. Wolff, *The Schrödinger Functional — a Renormalizable Probe for Non-Abelian Gauge Theories*, in preparation

- [31] U. Wolff, Nucl. Phys. B265 [FS15] (1986) 506 and 537
- [32] H. G. Evertz, M. Hasenbusch, M. Marcu, K. Pinn and S. Solomon, Nucl.Phys. B (Proc. Suppl.) 20 (1991) 80
- [33] M. Hasenbusch and S. Meyer, Phys. Rev. Lett. 66 (1991) 530
- [34] D. Kandel, R. Ben-Av and E. Domany, Phys. Rev. Lett. 65 (1990) 941
- [35] J.-S. Wang, R. H. Swendsen and Kotecký, Phys. Rev. Lett. 63 (1989) 109
- [36] R. Ben-Av, D. Kandel, E. Katznelson and P. G. Lauwers, J. Stat. Phys. 58 (1990) 125
- [37] R. C. Brower and S. Huang, Phys. Rev. D62 (1990) 708
- [38] B. Bunk, *An Efficient Algorithm for Cluster Updates in $Z(2)$ Lattice Gauge Theory*, in the Proceedings of the Workshop on Fermion Algorithms, HLRZ, Jülich 1991, eds. H. J. Herrmann and F. Karsch (World Scientific 1991)
- [39] For a new approach for $U(1)$ see:
 S. L. Adler and G. Bhanot, Phys. Rev. Lett. 66 (1991) 1807;
 S. L. Adler, G. Bhanot, T. Lippert, K. Schilling and P. Ueberholz,
Defeating Critical Slowing Down for Abelian Gauge Dynamics
 Princeton preprint IASSNS-HEP-91/39
- [40] M. Hasenbusch, S. Meyer and G. Mack,
 Nucl.Phys. B (Proc. Suppl.) 20 (1991) 110;
 M. Hasenbusch and S. Meyer, Phys. Rev. Lett. 68 (1992) 435;
Testing Accelerated Algorithms in the Lattice CP^3 Model, Kaiserslautern
 preprint Th 91-23 (1991)