

# Kinematics of Multigrid Monte Carlo

Martin Grabenstein<sup>1</sup> and Klaus Pinn<sup>2</sup>

Received August 10, 1992; final December 8, 1992

---

We study the kinematics of multigrid Monte Carlo algorithms by means of acceptance rates for nonlocal Metropolis update proposals. An approximation formula for acceptance rates is derived. We present a comparison of different coarse-to-fine interpolation schemes in free field theory, where the formula is exact. The predictions of the approximation formula for several interacting models are well confirmed by Monte Carlo simulations. The following rule is found: For a critical model with fundamental Hamiltonian  $\mathcal{H}(\phi)$ , the absence of critical slowing down can only be expected if the expansion of  $\langle \mathcal{H}(\phi + \psi) \rangle$  in terms of the shift  $\psi$  contains no relevant (mass) term. We also introduce a multigrid update procedure for non-abelian lattice gauge theory and study the acceptance rates for gauge group  $SU(2)$  in four dimensions.

---

**KEY WORDS:** Computer simulations; critical slowing down; multigrid Monte Carlo algorithms; acceptance rates; spin models; lattice gauge theory.

## 1. INTRODUCTION

Monte Carlo simulations have become an important tool for the study of critical phenomena in statistical mechanics<sup>(1)</sup> and in Euclidean quantum field theory on the lattice.<sup>(2)</sup> However, the method has limitations. In the vicinity of a critical point the phenomenon of critical slowing down (CSD) is a serious problem: for conventional local algorithms the autocorrelation time—that is, roughly speaking, the time used to generate a new “statistically independent” configuration—grows rapidly as the system approaches criticality. More precisely, the autocorrelation time  $\tau$  behaves like  $\tau \sim \xi^z$ , where  $\xi$  denotes the correlation length, and  $z$  is the dynamical critical exponent. For conventional local algorithms,  $z \approx 2$ . Thus, when the critical point is approached, there is a dramatic increase of computer time needed

---

<sup>1</sup> II. Institut für Theoretische Physik, Universität Hamburg, W-2000 Hamburg 50, Germany.

<sup>2</sup> Institut für Theoretische Physik I, Universität Münster, W-4400 Münster, Germany.

to calculate observables to a given accuracy. It is therefore important to devise new Monte Carlo algorithms that have reduced CSD.

For accelerated local algorithms such as overrelaxation or the optimized hybrid Monte Carlo algorithm, one can sometimes achieve  $z \approx 1$ <sup>(3-5)</sup> (see ref. 4 for a recent review). In an attempt to overcome the problem of CSD completely (in the sense of  $z \approx 0$ ), various nonlocal Monte Carlo algorithms have been developed.

Cluster algorithms<sup>(6)</sup> are successful in overcoming CSD for a large class of models. The alternative is multigrid Monte Carlo.<sup>(7-9)</sup> In this paper, every algorithm that updates stochastic variables on a hierarchy of length scales is called a multigrid Monte Carlo algorithm. There are models where no successful cluster algorithms have been found, whereas multigrid Monte Carlo algorithms work.<sup>(10,11)</sup>

Presently, the only generally applicable method to study algorithms for interacting models is by numerical experiment. For some models experiments show that the dynamical critical exponent  $z$  can be substantially reduced by a multigrid algorithm.<sup>(11-14)</sup> For other models, still  $z \approx 2$  is found.<sup>(15,16)</sup> An improved theoretical understanding of multigrid Monte Carlo algorithms is therefore desirable.

We present a method that can help to judge which algorithms will have a chance to overcome CSD in simulations of a given model *before* performing the simulation: We study the kinematics of multigrid Monte Carlo algorithms.<sup>3</sup> By kinematics we mean the study of the scale (block size) dependence of the Metropolis acceptance rates for nonlocal update proposals. We do not address the more difficult problem of analytically investigating the dynamical critical behavior of the stochastic processes involved. Our analysis is nonetheless of relevance because sufficiently high acceptance rates are necessary for multigrid Monte Carlo procedures to overcome CSD.

We derive an approximation formula for the block size dependence of acceptance rates for nonlocal Metropolis updates. The influence of the coarse-to-fine interpolation kernel (shape function) on the kinematics in free field theory, where the formula is exact, is investigated in detail.

The formula is then applied to several interacting models. It turns out to be a very good approximation. We find necessary criteria for a given multigrid algorithm to eliminate CSD: For a critical model with a fundamental Hamiltonian  $\mathcal{H}(\phi)$ , absence of CSD can only be expected if the expansion of  $\langle \mathcal{H}(\phi + \psi) \rangle$  in terms of the shift  $\psi$  contains no relevant term (mass term).

After introducing the study of the kinematics of multigrid Monte

<sup>3</sup> Parts of this paper have been published in short form in ref. 17.

Carlo in the context of spin models, where the multigrid methods have already been developed, we demonstrate that our analysis is also useful for the design of new multigrid procedures.

There is an urgent demand for accelerated Monte Carlo algorithms in lattice gauge theory. The present state-of-the-art algorithm is overrelaxation<sup>(18)</sup> (see ref. 19 for a review). However, effort has also been spent in developing nonlocal algorithms for gauge theories. An efficient cluster algorithm was found for  $SU(2)$  gauge theory at finite temperature, but only in the special case  $N_t = 1$ .<sup>(20)</sup> For a recent cluster algorithm approach to  $U(1)$  gauge theory see ref. 21. Multigrid algorithms for  $U(1)$  gauge models were introduced and studied in two and four dimensions.<sup>(13,22)</sup> A different but related nonlocal updating scheme in the Abelian case is the multiscale method.<sup>(23)</sup>

In this paper, we propose a multigrid algorithm for non-Abelian gauge theory and analyze its kinematics. Our approximation formula turns out to be very reliable also in this case and allows for a prediction of acceptance rates for a large class of nonlocal updates.

This paper is organized as follows: In Section 2 we introduce multigrid Monte Carlo algorithms. Section 3 contains the derivation of our approximation formula for acceptance rates. Several coarse-to-fine-interpolation kernels are discussed in Section 4. In Section 5 the acceptance rates in free field theory are examined in detail. The kinematical analysis for the sine-Gordon,  $XY$ , and  $\phi^4$  models is presented in Section 6. In Section 7 we propose a multigrid procedure for non-Abelian gauge theories and analyze its kinematics. A summary is given in Section 8.

## 2. MULTIGRID MONTE CARLO ALGORITHMS

We consider lattice models with partition functions

$$Z = \int \prod_{x \in A_0} d\phi_x \exp[-\mathcal{H}(\phi)] \quad (1)$$

on cubic  $d$ -dimensional lattices  $A_0$  with periodic boundary conditions. The lattice spacing is set to one. We use dimensionless spin variables  $\phi_x$ . An example is single-component  $\phi^4$ -theory, defined by the Hamiltonian

$$\mathcal{H}(\phi) = \frac{1}{2} (\phi, -\Delta\phi) + \frac{m_o^2}{2} \sum_x \phi_x^2 + \frac{\lambda_o}{4!} \sum_x \phi_x^4 \quad (2)$$

where

$$(\phi, -\Delta\phi) = \sum_{\langle x, y \rangle} (\phi_x - \phi_y)^2 \quad (3)$$

The sum in Eq. (3) is over all nearest neighbor pairs in the lattice. (The definitions for lattice gauge theory will be introduced in Section 7.)

A standard algorithm to perform Monte Carlo simulations of a model of the type defined above is the local Metropolis algorithm: One visits in a regular or random order the sites of the lattice and performs the following steps: At site  $x_o$ , one proposes a shift

$$\phi_{x_o} \rightarrow \phi'_{x_o} = \phi_{x_o} + s \quad (4)$$

The configuration  $\{\phi_x\}$  remains unchanged for  $x \neq x_o$ . Here  $s$  is a random number selected according to an *a priori* distribution  $\rho(s)$  which is symmetric with respect to  $s \rightarrow -s$ . For example, one selects  $s$  with uniform probability from an interval  $[-\varepsilon, \varepsilon]$ . One then computes the change of the Hamiltonian

$$\Delta \mathcal{H} = \mathcal{H}(\phi') - \mathcal{H}(\phi) \quad (5)$$

Finally the proposed shift is accepted with probability  $\min[1, \exp(-\Delta \mathcal{H})]$ . Then one proceeds to the next site.

The local Metropolis algorithm suffers from CSD when the correlation length in the system becomes large: long-wavelength fluctuations cannot efficiently be generated by a sequence of local operations. It is therefore natural to study nonlocal generalizations of the update procedure defined above.

Consider the fundamental lattice  $A_0$  as divided in cubic blocks of size  $l^d$ . This defines a block lattice  $A_1$ . By iterating this procedure, one obtains a whole hierarchy of block lattices  $A_0, A_1, \dots, A_K$  with increasing lattice spacing. This hierarchy of lattices is called a multigrid.

Let us denote block lattice points in  $A_k$  by  $x'$ . Block spins  $\Phi_{x'}$  are defined on block lattices  $A_k$ . They are averages of the fundamental field  $\phi_x$  over blocks of side length  $L_B = l^k$ :

$$\Phi_{x'} = L_B^{(d-2)/2} L_B^{-d} \sum_{x \in x'} \phi_x \quad (6)$$

The sum is over all points  $x$  in the block  $x'$ . The  $L_B$ -dependent factor in front of the average comes from the fact that the corresponding dimensional block spins are measured in units of the block lattice spacing: A scalar field  $\phi(x)$  in  $d$  dimensions has canonical dimension  $(2-d)/2$ . Thus  $\phi(x) = a^{(2-d)/2} \phi'_x$ , where  $a$  denotes the fundamental lattice spacing. Now measure the dimensional block spin  $\Phi(x')$  in units of the block lattice spacing  $a'$ :  $\Phi(x') = a'^{(2-d)/2} \Phi'_{x'}$ , with  $a' = aL_B$ . If we average in a natural way  $\Phi(x') = L_B^{-d} \sum_{x \in x'} \phi(x)$  and return to dimensionless variables, we obtain Eq. (6).

A nonlocal change of the configuration  $\phi$  consists of a shift

$$\phi_x \rightarrow \phi_x + s\psi_x \quad (7)$$

$s$  is a real parameter, and the ‘‘coarse-to-fine interpolation kernel’’ (or shape function)  $\psi_x$  determines the shape of the nonlocal change.  $\psi$  is normalized according to

$$L_B^{-d} \sum_{x \in x'} \psi_x = L_B^{(2-d)/2} \delta_{x', x'_o} \quad (8)$$

Note that by the nonlocal change (7), the block spin is moved as  $\Phi_{x'} \rightarrow \Phi_{x'} + s$  for  $x' = x'_o$ , and remains unchanged on the other blocks. The simplest choice of the kernel  $\psi$  that obeys the constraint (8) is a piecewise constant kernel:  $\psi_x = L_B^{(2-d)/2}$  if  $x \in x'_o$ , and 0 otherwise. Other kernels are smooth and thus avoid large energy costs from the block boundaries. A systematic study of different kernels will be given in Section 4 below.

The  $s$ -dependent Metropolis acceptance rate for such proposals is given by

$$\Omega(s) = \langle \min[1, \exp(-\Delta\mathcal{H})] \rangle \quad (9)$$

Here,  $\langle (\cdot) \rangle$  denotes the expectation value in the system defined by Eq. (1). Furthermore,

$$\Delta\mathcal{H} = \mathcal{H}(\phi + s\psi) - \mathcal{H}(\phi) \quad (10)$$

$\Omega(s)$  can be interpreted as the acceptance rate for shifting block spins by an amount of  $s$ , averaged over a sequence of configurations generated by a Monte Carlo simulation. Note that  $\Omega(s)$  does not depend on the algorithm that we use to compute it.  $\Omega(s)$  is a useful quantity when one wants to know how efficiently updates with increasing nonlocality (i.e., increasing block size  $L_B$ ) can be performed. Of course, different choices of the kernel  $\psi$  result in different acceptance rates.

In actual Monte Carlo simulations,  $s$  is not fixed. In the same way as in the local Metropolis algorithm,  $s$  is a random number distributed according to some *a priori* probability density. If we choose  $s$  to be uniformly distributed on the interval  $[-\varepsilon, \varepsilon]$ , the integrated acceptance rate  $P_{\text{acc}}$  (as customarily measured in Monte Carlo simulations) is obtained by averaging  $\Omega(s)$  as follows:

$$P_{\text{acc}}(\varepsilon) = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} ds \Omega(s) \quad (11)$$

It turns out to be a good rule to adjust the maximum Metropolis step size  $\varepsilon$  such that  $P_{\text{acc}}(\varepsilon) \approx 50\%$ .

We consider every algorithm that updates stochastic variables on a hierarchy of length scales as a multigrid Monte Carlo algorithm. However, there are two different classes of multigrid algorithms: multigrid algorithms in a unigrid implementation and “explicit” multigrid algorithms.

In the unigrid formulation one considers nonlocal updates of the form (7). Updates on the various layers of the multigrid are formulated on the level of the finest lattice  $A_0$ . There is no explicit reference to block spin variables  $\Phi$  defined on coarser layers  $A_k$  with  $k > 0$ . In addition, a unigrid also refers to a computational scheme: Nonlocal updates are performed directly on the level of the finest grid  $A_0$  in practical simulations.

In contrast, the explicit multigrid formulation consists of explicitly calculating conditional Hamiltonians depending on the block spin variables  $\Phi$  on coarser layers  $A_k$ . This formulation is possible if the conditional Hamiltonians are of the same type as or similar to the fundamental Hamiltonian. Then, the conditional probabilities used for the updating on the  $k$ th layer can be computed without always going back to the finest level  $A_0$ . Therefore, an explicit multigrid implementation reduces the computational work on the coarser layers (see the work estimates below). At least in free field theory, an explicit multigrid implementation is possible using nine-point prolongation kernels in two dimensions and generalizations thereof in higher dimensions.<sup>(15,24)</sup> Generally, an explicit multigrid implementation for interacting models is only feasible in special cases with piecewise constant kernels.

An algorithm formulated in the explicit multigrid style can always be translated to the unigrid language (that is how we are going to use the unigrid formulation). The reverse is not true, since not all nonlocal changes of the fundamental field configuration can be interpreted as updates of a single block spin variable of an explicit multigrid. As an example, one can use overlapping blocks in the unigrid style by translating the fields by a randomly chosen distance.<sup>(11)</sup>

If we formulate our kinematical analysis in the unigrid language, we nevertheless can include all algorithms formulated in the explicit multigrid style.

The sequence of sweeps through the different layers  $A_k$  of the multigrid is organized in a periodic scheme called a cycle.<sup>(25)</sup> The simplest scheme is the V-cycle: The sequence of layers visited in turn is  $A_0, A_1, \dots, A_K, A_{K-1}, \dots, A_1$ . More general cycles are characterized by the cycle control parameter  $\gamma$ . The rule is that from an intermediate layer  $A_k$  one proceeds  $\gamma$  times to the next coarser layer  $A_{k+1}$  before going back to the next finer layer  $A_{k-1}$ . A cycle control parameter  $\gamma > 1$  samples coarser layers more often than finer layers. With  $\gamma = 1$  we obtain the V-cycle.  $\gamma = 2$  yields the W-cycle that is frequently used with piecewise constant kernels.

The computational work estimates for the different cycles are as follows<sup>(15)</sup>: The work for an explicit multigrid cycle is  $\sim L^d$  if  $\gamma < l^d$ , where  $L$  denotes the lattice size. The work for a unigrid cycle is  $\sim L^d \log L$  if  $\gamma = 1$ , and  $\sim L^{d+\log_l \gamma}$  if  $\gamma > 1$ . Here,  $l$  denotes the blocking factor used in the iterative definition of the block lattices.

If one wants the computational work in the unigrid style to not exceed (up to a logarithm) an amount proportional to the volume  $L^d$  of the lattice, one has to use a V-cycle. Simulations with  $\gamma > 1$  (e.g., a W-cycle) can only be performed in the explicit multigrid style.

### 3. AN APPROXIMATION FORMULA FOR $\Omega(s)$

In this section we shall derive an approximate formula for the quantity  $\Omega(s)$  defined in (9). We can write  $\Omega(s)$  as

$$\Omega(s) = \int du \min(1, e^{-u}) \int \frac{dp}{2\pi} e^{-ipu} \langle e^{ip \Delta \mathcal{H}} \rangle \quad (12)$$

Let us assume that the probability distribution of  $\Delta \mathcal{H}$  is approximately Gaussian. We parametrize this distribution as follows:

$$d \text{prob}(\Delta \mathcal{H}) \propto d \Delta \mathcal{H} \exp \left[ -\frac{1}{4h_2} (\Delta \mathcal{H} - h_1)^2 \right] \quad (13)$$

with  $h_1 = \langle \Delta \mathcal{H} \rangle$  and  $h_2 = \frac{1}{2}(\langle \Delta \mathcal{H}^2 \rangle - \langle \Delta \mathcal{H} \rangle^2)$ . Then

$$\langle \exp(ip \Delta \mathcal{H}) \rangle \approx \exp(ih_1 p - h_2 p^2) \quad (14)$$

The integrations in Eq. (12) can be performed exactly since there are only Gaussian integrals involved. The result is

$$\Omega(s) \approx \frac{1}{2} \left[ \text{erfc} \left( \frac{h_1}{2\sqrt{h_2}} \right) + \exp(h_2 - h_1) \text{erfc} \left( \frac{2h_2 - h_1}{2\sqrt{h_2}} \right) \right] \quad (15)$$

with  $\text{erfc}(x) = 2/\sqrt{\pi} \int_x^\infty dt \exp(-t^2)$ . We shall now exploit the translational invariance of the measure  $\mathcal{D}\phi = \mathcal{D}(\phi + s\psi)$  to show that the difference of  $h_1$  and  $h_2$  is of order  $s^4$ . The starting point is the observation that  $\langle \exp(-\Delta \mathcal{H}) \rangle = 1$ . This implies that

$$\left. \frac{\partial^n}{\partial s^n} \ln \langle \exp(-\Delta \mathcal{H}) \rangle \right|_{s=0} = \left. \frac{\partial^n}{\partial s^n} \sum_{m=1}^n \frac{1}{m!} \langle (-\Delta \mathcal{H})^m \rangle_c \right|_{s=0} = 0 \quad (16)$$

$\langle (\cdot) \rangle_c$  denotes the connected (truncated) expectation value. Note that there are no contributions in the sum for  $m > n$ . This follows from the fact

that  $\Delta\mathcal{H}$  is of order  $s$ . Consequently,  $(\Delta\mathcal{H})^m = O(s^m)$ , and all contributions with  $m > n$  vanish in the limit  $s \rightarrow 0$ . For  $n = 2$  we obtain the relation

$$\frac{\partial^2}{\partial s^2} (\langle -\Delta\mathcal{H} \rangle + \frac{1}{2} \langle \Delta\mathcal{H}^2 \rangle_c) \Big|_{s=0} = \frac{\partial^2}{\partial s^2} (-h_1 + h_2) \Big|_{s=0} = 0 \quad (17)$$

If we assume that  $h_1$  and  $h_2$  are even in  $s$  (which is the case if  $\mathcal{H}$  is even in  $\phi$ ), then Eq. (17) says that the difference of  $h_1$  and  $h_2$  is of order  $s^4$ .

We shall later demonstrate that the approximation  $h_1 \approx h_2$  is in practice very good, even for small block size. In this case the acceptance rate prediction simplifies further,

$$\Omega(s) \approx \text{erfc}[\frac{1}{2}(h_1)^{1/2}] \quad (18)$$

(For an analogous result in the context of hybrid Monte Carlo see ref. 26.)

For free massless field theory with Hamiltonian  $\mathcal{H}(\phi) = \frac{1}{2}(\phi, -\Delta\phi)$ , we get  $h_1 = h_2 = \frac{1}{2}\alpha s^2$  with  $\alpha = (\psi, -\Delta\psi)$ , and our approximation formula becomes *exact*:

$$\Omega(s) = \text{erfc} \left[ \left( \frac{\alpha}{8} \right)^{1/2} |s| \right] \quad (19)$$

Equation (19) can be checked directly by using  $\langle \exp(ip \Delta\mathcal{H}) \rangle = \exp(ih_1 p - h_2 p^2)$  in Eq. (12). This relation is exact in free field theory.

#### 4. COARSE-TO-FINE INTERPOLATION

In this section we discuss several choices of the coarse-to-fine interpolation kernels. In order to have a “fair” comparison, all kernels  $\psi$  will be normalized according to Eq. (8).

In free massless field theory, the quantity  $\alpha = (\psi, -\Delta\psi)$  characterizes the decrease of the acceptance rate  $\Omega(s)$  of Eq. (19) with increasing shift  $s$ . Therefore it is natural to minimize  $\alpha$  in order to maximize  $\Omega(s)$  for fixed  $s$ .

The optimal kernel  $\psi^{\text{exact}}$  from the point of view of acceptance rates can be defined as follows: minimize the quadratic form

$$\alpha = (\psi, -\Delta\psi) \quad (20)$$

under the constraints that the average of  $\psi$  over the “central block”  $x'_0$  is given by  $L_B^{(2-d)/2}$  and its average over blocks  $x' \neq x'_0$  vanishes:

$$L_B^{-d} \sum_{x \in x'} \psi_x = L_B^{(2-d)/2} \delta_{x', x'_0} \quad \text{for all } x' \in \mathcal{A}_k \quad (21)$$



This variational problem can be solved with the help of Fourier methods. The result is

$$\psi_x^{\text{exact}} = L_B^{(2+d)/2} \mathcal{A}_{x,x'_0} \quad (22)$$

where  $\mathcal{A}_{x,x'_0}$  denotes the Gawędzki–Kupiainen kernel (see, e.g., ref. 9). The use of this kernel leads to a complete decoupling of the different layers of the multigrid. This way of interpolating from a coarser block lattice  $A_k$  to the fine lattice  $A_0$  is well known in rigorous renormalization group theory.<sup>(27)</sup> It is interesting that considerations about optimizing acceptance rates in a stochastic multigrid procedure lead to the same choice of the interpolation kernel.

Because  $\psi^{\text{exact}}$  is nonvanishing on the whole lattice, it is not convenient for numerical purposes. For an attempt to change the block spin  $\Phi_{x'_0}$  on block  $x'_0$  one has contributions to the change of the Hamiltonian from all lattice points. Therefore the computational work for a single update is proportional to the volume.

We define a “truncated kernel”  $\psi^{\text{trunc}}$  by restricting the support of  $\psi$  on the block  $x'_0$  and its nearest neighbor blocks  $y'_0$ :

$$\psi_x^{\text{trunc}} = 0 \quad \text{if } x \notin x'_0 \text{ or } x \notin y'_0, \quad \text{where } y'_0 \text{ n.n. } x'_0 \quad (23)$$

In other words, the Laplacian in Eq. (20) is replaced by a Laplacian  $\Delta_D$  with Dirichlet boundary conditions on the boundary of the support of  $\psi$ . We again minimize  $\alpha = (\psi, -\Delta_D \psi)$  under the  $2d+1$  constraints that the average of  $\psi$  over the blocks  $x'_0$  and its nearest neighbor blocks is given. This minimization can be performed numerically by a relaxation procedure. In order to maintain the normalization condition, one always updates simultaneously two spins residing in the same block, keeping their sum fixed. The  $\psi^{\text{trunc}}$  kernels were used in a multigrid simulation of the  $\phi^4$  model in four dimensions.<sup>(28)</sup>

From a practical point of view, it is convenient to use kernels that have support on a single block  $x'_0$ , i.e.,

$$\psi_x = 0 \quad \text{if } x \notin x'_0 \quad (24)$$

We define a kernel  $\psi^{\text{min}}$  with this property by minimizing  $\alpha = (\psi, -\Delta_{D,x'_0} \psi)$  under the constraint that the average of  $\psi$  over the block  $x'_0$  is given. The Laplacian with Dirichlet boundary conditions on the boundary of  $x'_0$  is defined as follows:

$$(\Delta_{D,x'_0} \phi)_x = \left( -2d\phi_x + \sum_{\substack{y \text{ n.n. } x \\ y \in x'_0}} \phi_y \right) \quad \text{for } x \in x'_0 \quad (25)$$

$\psi^{\text{min}}$  can be calculated using an orthonormal set of eigenfunctions of  $\Delta_{D,x'_0}$ .

We shall now discuss other kernels with support on the block that are frequently used in the literature.

*Piecewise constant interpolation:*

$$\psi_x^{\text{const}} = \begin{cases} L_B^{(2-d)/2} & \text{for } x \in x'_o \\ 0 & \text{for } x \notin x'_o \end{cases} \quad (26)$$

This kernel has the advantage that for many models the conditional Hamiltonians used for updating on coarse grids are of the same type as or are similar to the fundamental Hamiltonian. This means that the conditional probabilities used for the updating on the  $k$ th layer can be computed without always going back to the finest level  $A_o$ . Therefore, an explicit multigrid implementation with a W-cycle can be used.

*Piecewise linear interpolation:* We consider the block

$$x'_o = \{x \mid x^\mu \in \{1, 2, 3, \dots, L_B\}, \mu = 1, \dots, d\} \quad (27)$$

The kernels for other blocks are simply obtained by translation. For  $L_B$  even,  $\psi_x^{\text{linear}}$  is given by

$$\psi_x^{\text{linear}} = \mathcal{N} \prod_{\mu=1}^d \left\{ \frac{L_B+1}{2} - \left| x^\mu - \frac{L_B+1}{2} \right| \right\} \mathbb{A} \quad \text{for } x \in x'_o \quad (28)$$

$\mathcal{N}$  is a normalization constant.

*Ground-state projection kernels:*  $\psi^{\text{sine}}$  is the eigenfunction corresponding to the lowest eigenvalue of the negative Laplacian with Dirichlet boundary conditions  $-\Delta_{D, x'_o}$ :

$$\psi_x^{\text{sine}} = \begin{cases} \mathcal{N} \prod_{\mu=1}^d \sin\left(\frac{\pi}{L_B+1} x^\mu\right) & \text{for } x \in x'_o \\ 0 & \text{for } x \notin x'_o \end{cases} \quad (29)$$

Again,  $\mathcal{N}$  denotes a normalization constant. Note that this kernel is different from  $\psi^{\text{min}}$ . A generalization of this kernel was introduced for scalar fields in the background of nonabelian gauge fields.<sup>(29)</sup>

The results for the quantities  $\alpha = (\psi, -\Delta\psi)$  for different kernels in two dimensions are presented in Table I. We used a  $512^2$  lattice ( $\psi^{\text{exact}}$  depends on the lattice size). The different kernels are ordered according to increasing value of  $\alpha$ .

The values of  $\alpha^{\text{exact}}$  and  $\alpha^{\text{trunc}}$  are close together. This shows that the truncation of the support of  $\psi$  to the block and its nearest neighbor blocks is a good approximation to  $\psi^{\text{exact}}$  (in the sense of acceptance rates). The

Table I. Results for  $\alpha = (\Psi, -\Delta\Psi)$  in Two Dimensions,  $512^2$  Lattice

Kernel	$L_B=2$	$L_B=4$	$L_B=8$	$L_B=16$	$L_B=32$	$L_B=64$	$L_B=128$	$L_B=256$
Exact	6.899	8.902	9.705	9.941	10.00	10.02	10.18	13.11
Trunc	7.000	9.405	10.73	11.38	11.69	11.84	11.92	—
Min	8.000	13.24	18.48	22.58	25.23	26.76	27.59	28.02
Sine	8.000	13.62	19.34	23.78	26.62	28.25	29.13	29.58
Linear	8.000	15.80	24.58	31.84	36.68	39.51	41.05	41.84
Const	8.000	16.00	32.00	64.00	128.0	256.0	512.0	1024

value of  $\alpha^{\text{exact}}$  for  $L_B=256$  is remarkably higher than on smaller blocks. This is a finite-size effect because the block lattice consists only of  $2^2$  points. Since the nearest neighbors overlap on a  $2^2$  lattice, no result for  $\alpha^{\text{trunc}}$  is quoted for  $L_B=256$ . The values of  $\alpha$  for the smooth kernels with support on the blocks  $\psi^{\text{min}}$ ,  $\psi^{\text{sine}}$ , and  $\psi^{\text{linear}}$  are of the same magnitude. We can see that  $\psi^{\text{sine}}$  is almost as good as the optimal  $\psi = \psi^{\text{min}}$ .

The results for different kernels in four dimensions are presented in Table II. Here we used a  $64^4$  lattice. In principle, the  $\alpha$ 's behave as in two dimensions. The values of  $\alpha^{\text{linear}}$  for small blocks are higher than  $\alpha^{\text{const}}$ . The pyramids of the piecewise linear kernels have many edges in four dimensions, which lead to high costs in the kinetic energy.

The  $L_B$  dependence of the  $\alpha$ 's in  $d$  dimensions is

$$\begin{aligned} \alpha &= 2dL_B && \text{for piecewise constant kernels} \\ \alpha &\xrightarrow[L_B \gg 1]{} \text{const} && \text{for smooth kernels} \end{aligned} \quad (30)$$

As an example, the expression for  $\alpha^{\text{sine}}$  in  $d$  dimensions is

$$\alpha^{\text{sine}} = L_B^{2+d}(L_B+1)^d 2^{d+2} d \sin^{4d+2} \left[ \frac{\pi}{2(L_B+1)} \right] \sin^{-2d} \left( \frac{\pi}{L_B+1} \right) \quad (31)$$

Table II. Results for  $\alpha = (\Psi, -\Delta\Psi)$  in Four Dimensions,  $64^4$  Lattice

Kernel	$L_B=2$	$L_B=4$	$L_B=8$	$L_B=16$	$L_B=32$
Exact	14.48	20.38	23.48	24.71	30.62
Trunc	14.67	21.61	26.54	29.26	—
Min	16.00	27.72	41.56	54.18	63.33
Sine	16.00	30.37	48.46	64.85	76.44
Linear	16.00	39.02	70.78	101.0	122.9
Const	16.00	32.00	64.00	128.0	256.0

For large block sizes we find

$$\alpha^{\text{sine}} \xrightarrow{L_B \gg 1} d \frac{\pi^{2d+2}}{2^{3d}} = \text{const} \quad (32)$$

From Table I we observe that in two dimensions  $\alpha$  becomes almost independent of  $L_B$  for the smooth kernels if the block size is larger than 16. In four dimensions (Table II), we find  $\alpha(L_B) \sim \text{const}$  only for  $\alpha^{\text{exact}}$ . The other  $\alpha$ 's for the smooth kernels have not become independent of  $L_B$  for the block sizes studied.

## 5. ACCEPTANCE RATES IN FREE FIELD THEORY

Recall that we have  $\Omega(s) = \text{erfc}[(\alpha/8)^{1/2} |s|]$  in massless free field theory.  $\Omega(s)$  is only a function of the product  $\alpha s^2$ . In order to keep  $\Omega(s)$  fixed when the block size  $L_B$  increases, we have to rescale the changes  $s$  like  $\alpha(L_B)^{-1/2}$ . As a consequence, to maintain a constant acceptance rate in massless free field theory,  $s$  has to be scaled down like  $L_B^{-1/2}$  for piecewise constant kernels, whereas for smooth kernels the acceptance rates for large  $L_B$  do not depend on the block size.

Note that this behavior of the acceptance rates for large  $L_B$  is not yet reached in four dimensions for the block sizes studied (except for  $\psi^{\text{exact}}$ ). See also the discussion of the Metropolis step size below. At least for free field theory, the disadvantage of the piecewise constant kernels can be compensated for by using a W-cycle instead of a V-cycle. Smooth kernels can be used only in V-cycle algorithms. Exceptions are nine-point prolongation kernels in two dimensions and generalizations thereof in higher dimensions. They can also be used with a W-cycle, at least in free field theory (cf. Section 2).

We now illustrate what this rescaling of  $s$  means for the Metropolis step size  $\varepsilon$  in an actual multigrid Monte Carlo simulation. Look at the integrated acceptance probability defined in Eq. (11). If we insert the exact result (19) for massless free field theory, we get

$$P_{\text{acc}}(\varepsilon) = \text{erfc} \left[ \left( \frac{\alpha}{8} \right)^{1/2} \varepsilon \right] + \frac{1}{(\pi\alpha/8)^{1/2} \varepsilon} [1 - e^{-(\alpha/8) \varepsilon^2}] \quad (33)$$

$P_{\text{acc}}$  is only a function of the product  $\alpha \varepsilon^2$ . In order to keep  $P_{\text{acc}}$  fixed (to, e.g., 50%), we have to rescale  $\varepsilon(L_B)$  like  $\alpha(L_B)^{-1/2}$ , exactly in the same way as we had to rescale  $s$  to keep  $\Omega(s)$  fixed. This  $L_B$  dependence is plotted in Fig. 1 for two dimensions and in Fig. 2 for four dimensions.

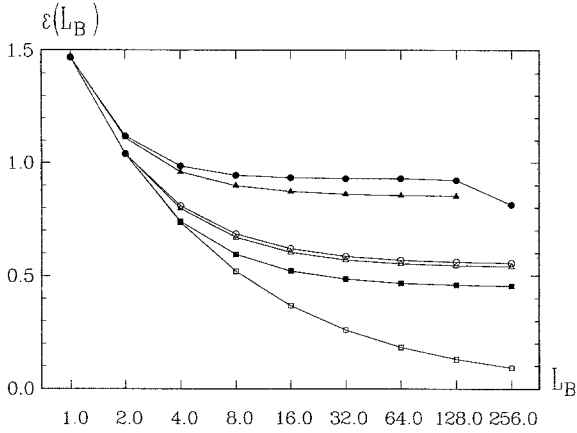


Fig. 1. Metropolis step sizes  $\varepsilon(L_B)$  for massless free field theory in two dimensions,  $512^2$  lattice.  $\varepsilon(L_B)$  is chosen in such a way that always  $P_{\text{acc}} = 0.5$  holds. Symbols: full circles,  $\psi^{\text{exact}}$ ; full triangles,  $\psi^{\text{trunc}}$ ; empty circles,  $\psi^{\text{min}}$ ; empty triangles,  $\psi^{\text{sinc}}$ ; full squares,  $\psi^{\text{linear}}$ ; empty squares,  $\psi^{\text{const}}$ . Lines are only drawn to guide the eye.

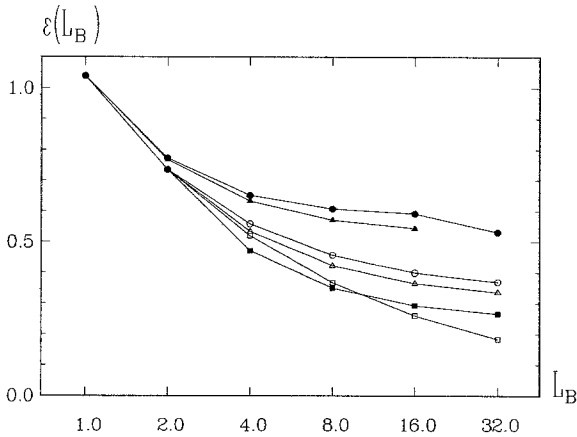


Fig. 2. Metropolis step sizes  $\varepsilon(L_B)$  for massless free field theory in four dimensions,  $64^4$  lattice.  $\varepsilon(L_B)$  is chosen in such a way that always  $P_{\text{acc}} = 0.5$  holds. Symbols: full circles,  $\psi^{\text{exact}}$ ; full triangles,  $\psi^{\text{trunc}}$ ; empty circles,  $\psi^{\text{min}}$ ; empty triangles,  $\psi^{\text{sinc}}$ ; full squares,  $\psi^{\text{linear}}$ ; empty squares,  $\psi^{\text{const}}$ . Lines are only drawn to guide the eye.

We now discuss massive free field theory with Hamiltonian  $\mathcal{H}(\phi) = \frac{1}{2}(\phi, [-\Delta + m^2]\phi)$ . We find  $h_1 = \frac{1}{2}\alpha_m s^2$ , with  $\alpha_m$  given by

$$\alpha_m = (\psi, [-\Delta + m^2]\psi) = \alpha + m^2 \sum_{x \in \mathcal{A}_0} \psi_x^2 \quad (34)$$

Therefore the exact result is  $\Omega(s) = \text{erfc}[(\alpha_m/8)^{1/2} |s|]$ . The term  $\sum_x \psi_x^2$  scales  $\sim L_B^2$  in arbitrary dimensions. For piecewise constant kernels

$$\sum_{x \in \mathcal{A}_0} (\psi_x^{\text{const}})^2 = L_B^2 \quad (35)$$

For  $\psi^{\text{sine}}$  kernels we find

$$\sum_{x \in \mathcal{A}_0} (\psi_x^{\text{sine}})^2 = 2^d L_B^{2+d} (L_B + 1)^d \sin^{4d} \left[ \frac{\pi}{2(L_B + 1)} \right] \sin^{-2d} \left( \frac{\pi}{L_B + 1} \right) \quad (36)$$

and for large block sizes

$$\sum_{x \in \mathcal{A}_0} (\psi_x^{\text{sine}})^2 \xrightarrow{L_B \gg 1} \frac{\pi^{2d}}{2^{3d}} L_B^2 \quad (37)$$

If the block size  $L_B$  is smaller than the correlation length  $\xi = 1/m$ ,  $h_1$  is still dominated by the kinetic term  $s^2(\psi, -\Delta\psi)$ , and the discussion is the same as in the massless case.

As soon as the block size  $L_B$  becomes larger than  $\xi$ ,  $h_1$  is dominated by the mass term  $s^2 m^2 \sum_x \psi_x^2 \sim s^2 L_B^2$ , and  $s$  has to be rescaled like  $s \sim L_B^{-1}$  in order to maintain constant acceptance rates. Of course this is a dramatic decrease for large block sizes compared to  $s \sim \text{const}$  (using smooth kernels) in the massless case. Block spins on large blocks are essentially “frozen.” But this is not dangerous for the performance of the algorithm in massive free field theory: The effective probability distribution for the block spins  $\Phi$  is given by  $\exp[-\mathcal{H}_{\text{eff}}(\Phi)]$ , where  $\mathcal{H}_{\text{eff}}(\Phi)$  denotes the effective Hamiltonian in the sense of the block spin renormalization group (for a recent review see ref. 30). The physical fluctuations of the block spins are dictated by an effective mass term

$$m_{\text{eff}}^2 \sum_{x' \in \mathcal{A}_k} \Phi_{x'}^2 \quad \text{with} \quad m_{\text{eff}}^2 \sim m^2 L_B^2 \quad (38)$$

Thus, the algorithmic fluctuations (described by the mass term  $m^2 \sum_x \psi_x^2 \sim m^2 L_B^2$ ) and the physical fluctuations (described by the effective mass  $\sim m^2 L_B^2$ ) behave in a similar way, and the multigrid algorithm is able to create fluctuations just of the size that is needed by the physics of the

model. Moreover, there is no need to do updates at length scales larger than  $\xi$  in order to beat CSD.

In this sense, the discussed algorithmic mass term  $m^2 \sum_x \psi_x^2$  is well behaved for free field theory, since it decreases with the physical mass in the vicinity of the critical point. As we shall see in Section 6, for interacting models close to criticality, a different scenario is possible. There, it can happen that an algorithmic “mass term”  $\sim \sum_x \psi_x^2$  persists, whereas the renormalized mass vanishes. If this happens, the multigrid algorithm is not able to produce the large critical fluctuations required by the physics, and we *cannot* expect that CSD will be eliminated.

The  $L_B$  dependence of the term  $\sum_x \psi_x^4$  will also be needed in the study of the  $\phi^4$  theory in Section 6 below. In  $d$  dimensions such a term scales  $\sim L_B^{4-d}$ : For piecewise constant kernels

$$\sum_{x \in A_0} (\psi_x^{\text{const}})^4 = L_B^{4-d} \tag{39}$$

whereas using  $\psi^{\text{sine}}$  kernels, we find

$$\sum_{x \in A_0} (\psi_x^{\text{const}})^4 = 6^d L_B^{4+2d} (L_B + 1)^d \sin^{8d} \left[ \frac{\pi}{2(L_B + 1)} \right] \sin^{-4d} \left( \frac{\pi}{L_B + 1} \right) \tag{40}$$

In the limit of large block sizes this term behaves like

$$\sum_{x \in A_0} (\psi_x^{\text{sine}})^4 \xrightarrow{L_B \gg 1} \left( \frac{3\pi^4}{128} \right)^d L_B^{4-d} \tag{41}$$

In order to summarize the different large- $L_B$  behaviors of local operators in the kernel  $\psi$  discussed here, let us introduce the *degree of relevance* in the sense of the perturbative renormalization group: The (superficial) degree  $r$  of relevance of a local operator in  $\psi$  which is a polynomial of  $m$  scalar fields with  $n$  derivatives is defined by  $r = d + m(2 - d)/2 - n$ . This definition is valid for smooth kernels. For large  $L_B$ , an operator with degree of relevance  $r$  behaves like  $L_B^r$ . An operator is called relevant if  $r > 0$ . As we have seen in the examples above, a mass term has  $r = 2$ , and a  $\psi^4$  term has  $r = 4 - d$ . A kinetic term  $\alpha = (\psi, -\Delta\psi)$  has  $r = 0$  for smooth kernels.

The only difference for piecewise constant kernels is that the kinetic term behaves like  $\alpha = (\psi, -\Delta\psi) \propto L_B$ .

## 6. ACCEPTANCE RATES FOR INTERACTING MODELS

In this section, we shall apply formula (18) in the discussion of multigrid procedures for three different spin models in two dimensions: the

sine-Gordon model, the  $XY$  model, and the single-component  $\phi^4$  theory. The scale dependence of acceptance rates for interacting models will be compared with the behavior in free field theory, where CSD is known to be eliminated by a multigrid algorithm.

### 6.1. Two-Dimensional Sine-Gordon Model

The two-dimensional sine-Gordon model is defined by the Hamiltonian

$$\mathcal{H}(\phi) = \frac{1}{2\beta} (\phi, -\Delta\phi) - \zeta \sum_x \cos \phi_x \quad (42)$$

The model undergoes a Kosterlitz–Thouless phase transition at  $\beta_c$ . In the limit of vanishing fugacity  $\zeta$ , the location of the critical  $\beta$  is exactly known:  $\beta_c \rightarrow 8\pi$  for  $\zeta \rightarrow 0$ . For  $\beta > \beta_c$ , the model is in the massless phase, and the flow of the effective Hamiltonian (in the sense of the block spin renormalization group) converges to that of a massless free field theory: the long-distance behavior of the theory is that of a Gaussian model. Since multigrid algorithms have proven to be efficient in generating long-wavelength Gaussian modes, one might naively conclude that multigrid should be the right method to fight CSD in the simulation of the sine-Gordon model in the massless phase. But this is not so. For  $h_1$  we find the expression

$$h_1 = \frac{\alpha}{2\beta} s^2 + \zeta C \sum_x [1 - \cos(s\psi_x)] \quad (43)$$

with  $C = \langle \cos \phi_x \rangle$ . Recall that  $h_1$  is the quantity that determines the acceptance rates  $\Omega(s)$ :

$$\Omega(s) \approx \operatorname{erfc}[\frac{1}{2}(h_1)^{1/2}] \quad (44)$$

The essential point is that the second term in (43) is proportional to the block volume  $L_B^2$  for piecewise constant *and* for smooth kernels (see the discussion in Section 5). This can be checked for small  $s$  by expanding in  $s$ . One therefore has to face a dramatic decrease of acceptance when the blocks become large, even for small fugacity  $\zeta$ . A constant acceptance rate is achieved only when the proposed steps are scaled down like  $L_B^{-1}$ . It is therefore unlikely that any multigrid algorithm—based on nonlocal updates of the type discussed in this paper—will be successful for this model.

We demonstrate the validity of formula (18) (using a Monte Carlo estimate for  $C$ ) by comparing with Monte Carlo results at  $\beta = 39.478$  and



$\zeta = 1$ . This point is in the massless phase, where the correlation length  $p$  is of the order of the lattice size  $L$ . In Fig. 3 we show both the numerical and analytical results for  $\Omega(s)$  for  $L_B = 4, 8, 16,$  and  $32$  on lattices of size  $16^2, 32^2, 64^2,$  and  $128^2,$  respectively.

We tested the precision of our approximation formula for piecewise constant kernels only. However, we have no doubts that the quality of the approximation is also very good for other shape functions  $\psi$ .

### 6.2. Two-Dimensional XY Model

We now discuss the two-dimensional XY model, defined by the partition function

$$Z = \int \prod_x d\theta_x \exp \left[ \beta \sum_{\langle x, y \rangle} \cos(\theta_x - \theta_y) \right] \tag{45}$$

The sum is over all unordered pairs of nearest neighbors in the lattice. Like the sine-Gordon model, the XY model has a massless (spin wave) phase for  $\beta > \beta_c$  and a massive phase for  $\beta < \beta_c$ . The best available estimate for the critical coupling is  $\beta_c = 1.1197(5)$ .<sup>(31)</sup>

Nonlocal updates are defined by

$$\theta_x \rightarrow \theta_x + s\psi_x \tag{46}$$

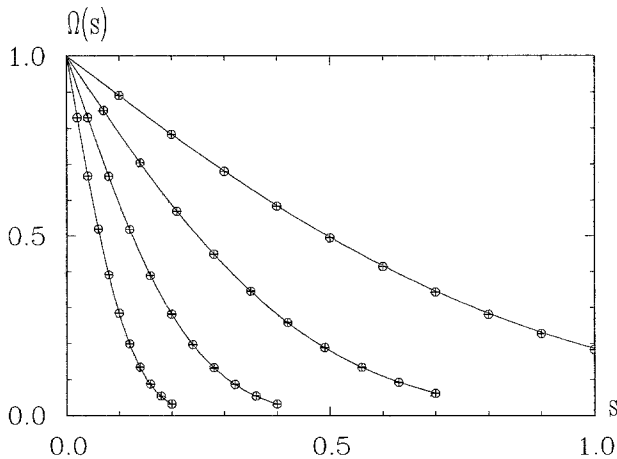


Fig. 3.  $\Omega(s)$  for piecewise constant kernels in the two-dimensional sine-Gordon model,  $\beta = 39.478, \zeta = 1$ . From top to bottom:  $L_B = 4, 8, 16,$  and  $32$  on a  $16^2, 32^2, 64^2,$  and  $128^2$  lattice, respectively. Points with error bars: Monte Carlo results; lines: analytical results.

with  $\psi$  obeying again the normalization condition (8). To define a (linear) block spin, we rewrite the partition function (45) in terms of two-component unit vector spin variables  $s_x$ :

$$Z = \int \prod_x [d^2 s_x \delta(s_x^2 - 1)] \exp\left(\beta \sum_{\langle x, y \rangle} s_x \cdot s_y\right) \quad (47)$$

The block spins  $S_x$  are then defined as block averages of the unit vectors  $s_x$ .

Note that the proposal (46) changes the block spin by an amount  $\approx s$  only when the spins inside the block are sufficiently aligned. This will be the case in the spin-wave phase for large enough  $\beta$ . For smaller  $\beta$ , the correct (or "fair") normalization of the kernels  $\psi$  is a subtle point. We believe, however, that our argument is not affected by this in a qualitative way.

The relevant quantity  $h_1$  is given by

$$h_1 = \beta E \sum_{\langle x, y \rangle} \{1 - \cos[s(\psi_x - \psi_y)]\} \quad (48)$$

with  $E = \langle \cos(\Theta_x - \Theta_y) \rangle$ ;  $x$  and  $y$  nearest neighbors. For piecewise constant kernels,  $h_1$  is proportional to  $L_B$ . For smooth kernels  $h_1$  will become independent of  $L_B$  for large enough blocks. For small  $s$ ,

$$h_1 \approx \frac{1}{2} s^2 \beta E \sum_{\langle x, y \rangle} (\psi_x - \psi_y)^2 = \frac{1}{2} s^2 \beta E \alpha \quad (49)$$

As above,  $\alpha = \langle (\psi_x - \psi_y)^2 \rangle$ . This quantity becomes nearly independent of  $L_B$  already for  $L_B$  larger than 16 (cf. Section 4).

From the point of view of acceptance rates the XY model therefore behaves like massless free field theory. A dynamical critical exponent  $z$  consistent with zero was observed in the massless phase.<sup>(32)</sup> The failure of multigrid Monte Carlo in the massive phase ( $z \approx 1.4$  for piecewise constant kernels<sup>(32)</sup>) is an example of the fact that good acceptance rates are not sufficient to overcome CSD.

We again checked the accuracy of formula (18) by comparing with Monte Carlo results at  $\beta = 1.2$  (which is in the massless phase, where the correlation length  $\xi$  is of the order of the lattice size  $L$ ). The only numerical input for the analytical formula was the link expectation value  $E$ . The results are displayed in Fig. 4.

One can make a similar discussion for the  $O(N)$  nonlinear  $\sigma$  model with  $N > 2$ , leading to the same prediction for the scale dependence of the acceptance rates. This behavior was already observed in multigrid Monte Carlo simulations of the  $O(3)$  nonlinear  $\sigma$  model in two dimensions with smooth and piecewise constant kernels.<sup>(12)</sup>

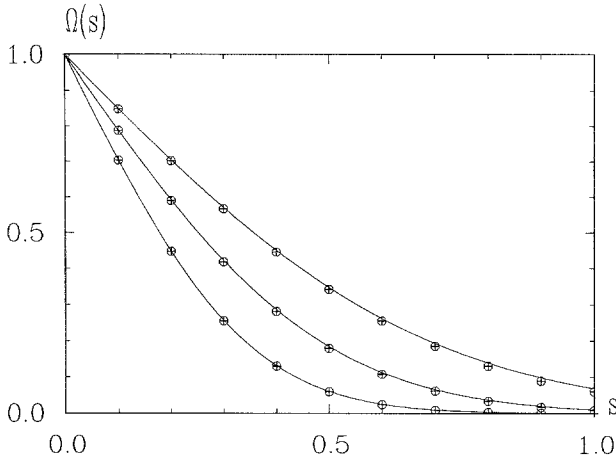


Fig. 4.  $\Omega(s)$  for piecewise constant kernels in the two-dimensional  $XY$  model,  $\beta = 1.2$ . From top to bottom:  $L_B = 4, 8,$  and  $16$  on a  $16^2, 32^2,$  and  $64^2$  lattice, respectively. Points with error bars: Monte Carlo results; lines: analytical results.

### 6.3. Two-Dimensional $\phi^4$ Theory

Let us now turn to single-component  $d$ -dimensional  $\phi^4$  theory, defined by the Hamiltonian

$$\mathcal{H}(\phi) = \frac{1}{2} (\phi, -\Delta\phi) + \frac{m_o^2}{2} \sum_x \phi_x^2 + \frac{\lambda_o}{4!} \sum_x \phi_x^4 \tag{50}$$

For  $h_1$  one finds

$$h_1 = s^2 \left[ \frac{1}{2} \alpha + \left( \frac{m_o^2}{2} + \frac{\lambda_o}{4} P \right) \sum_x \psi_x^2 \right] + s^4 \frac{\lambda_o}{4!} \sum_x \psi_x^4 \tag{51}$$

where  $P = \langle \phi_x^2 \rangle$ . We have used the fact that expectation values of operators which are odd in  $\phi$  vanish on finite lattices. Recall that  $\sum_x \psi_x^2$  increases with  $L_B^2$ , independent of  $d$ , whereas  $\sum_x \psi_x^4$  scales like  $L_B^{4-d}$  for smooth and for piecewise constant kernels (see the discussion of the different kernels in Section 5). We conclude that also in this model we have to face rapidly decreasing acceptance rates when the blocks become large. As in the case of the sine-Gordon model,  $s$  has to be rescaled like  $L_B^{-1}$  in order to maintain constant acceptance rates.

Therefore there is little hope that multigrid algorithms of the type considered here can overcome CSD in the one-component  $\phi^4$  model. In

numerical simulations of two-dimensional  $\phi^4$  theory, a dynamical critical behavior is found that is consistent with  $z \approx 2$  for piecewise constant and for smooth kernels.<sup>(8,15,16)</sup> In four dimensions, there is no definite estimate for  $z$ .<sup>(28)</sup>

Figure 5 shows a comparison of Monte Carlo results for two-dimensional  $\phi^4$  theory with the theoretical prediction based on the numerical evaluation of  $P$ . The simulations were done in the symmetric phase at  $m_o^2 = -0.56$  and  $\lambda_o = 2.4$ . The correlation length at this point is  $\xi \approx 15$ .<sup>(16)</sup>

#### 6.4. Summary of Section 6

Our approximation formula has proven to be quite precise. The results for three different models are consistent with the following rule:

Sufficiently high acceptance rates for a complete elimination of CSD can only be expected if  $h_1 = \langle \mathcal{H}(\phi + s\psi) - \mathcal{H}(\phi) \rangle$  contains no relevant operator in  $\psi$ .

As we have seen above, the typical candidate for a relevant operator in  $\psi$  is always an "algorithmic mass term" of the type  $\sim s^2 \sum_x \psi_x^2$  with degree of relevance  $r = 2$ .

This rule is formulated for smooth kernels. For piecewise constant kernels, it has to be modified. There, the kinetic term  $\alpha \propto L_B$  is relevant as well. At least in free field theory this disadvantage can be compensated for

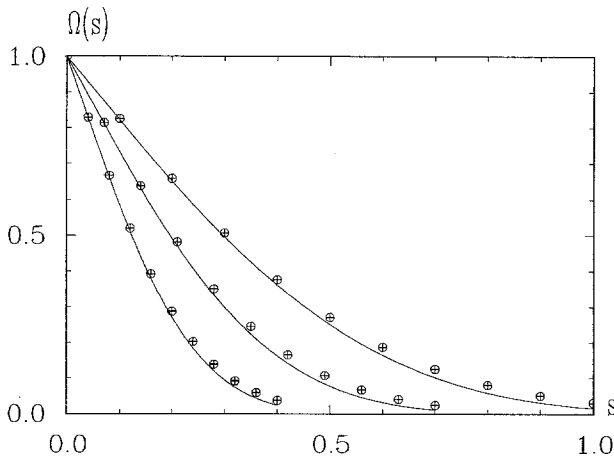


Fig. 5.  $\Omega(s)$  for piecewise constant kernels in the two-dimensional  $\phi^4$  theory,  $m_o^2 = -0.56$ ,  $\lambda_o = 2.4$ . From top to bottom:  $L_B = 4, 8$ , and  $16$  on a  $16^2, 32^2$ , and  $64^2$  lattice, respectively. Points with error bars: Monte Carlo results; lines: analytical results.

by using a W-cycle. Apart from this modification the rule carries over to the case of piecewise constant kernels.

In the following we are going to use our method of investigating the kinematics of a multigrid Monte Carlo algorithm for the development of new multigrid schemes. One is able to see *before* actually simulating with a new technique whether the algorithm under study will have a chance to overcome CSD. We focus on the problem of finding fast algorithms for lattice gauge fields.

## 7. A MULTIGRID PROCEDURE FOR LATTICE GAUGE FIELDS

In this section we propose a multigrid procedure for pure lattice gauge theory and study the behavior of acceptance rates with increasing block size  $L_B$ .

We consider partition functions

$$Z = \int \prod_{x,\mu} dU_{x,\mu} \exp[-\mathcal{H}(U)] \quad (52)$$

The link variables  $U_{x,\mu}$  take values in the gauge group  $U(1)$  or  $SU(N)$ , and  $dU$  denotes the corresponding invariant Haar measure. The standard Wilson action  $\mathcal{H}(U)$  is given by

$$\mathcal{H}(U) = \beta \sum_{\mathcal{P}} \left( 1 - \frac{1}{N} \text{Re Tr } U_{\mathcal{P}} \right) \quad (53)$$

The sum in (53) is over all plaquettes in the lattice. The  $U_{\mathcal{P}}$  are path-ordered products around plaquettes  $\mathcal{P}$ ,

$$U_{\mathcal{P}} = U_{x,\mu} U_{x+\hat{\mu},\nu} U_{x+\hat{\nu},\mu}^* U_{x,\nu}^* \quad (54)$$

$U^*$  denotes the Hermitian conjugate (=inverse) of  $U$ .

### 7.1. The Abelian Case

We now consider the case of gauge group  $U(1)$ . The link variables are parametrized with an angle  $-\pi \leq \theta_{x,\mu} < \pi$  through

$$U_{x,\mu} = \exp(i\theta_{x,\mu}) \quad (55)$$

Nonlocal updates can be defined as follows: One chooses a hypercubic block  $x'_o$  of size  $L_B^d$  and a direction  $\tau$  with  $1 \leq \tau \leq d$ . During the update,  $\tau$

will be kept fixed. All the link variables  $U_{x,\tau}$  attached to sites  $x$  inside the block  $x'_o$  are proposed to be changed simultaneously:

$$U_{x,\tau} \rightarrow \exp(is\psi_x) U_{x,\tau} \quad (56)$$

Again,  $\psi$  denotes an interpolation kernel as introduced in Section 4. This updating scheme was introduced and studied in two dimensions in ref. 13 and also in four dimensions in ref. 22, using piecewise constant kernels. Of course, one can use all versions of smooth kernels as well, with their support not necessarily restricted to the block  $x'_o$ .

Let us now study the acceptance rates for these update proposals with the help of formula (18). We consider general kernels  $\psi$ . For  $h_1 = \langle \Delta \mathcal{H} \rangle$  we find

$$h_1 = \beta P \sum_{x \in A_0} \sum_{\mu \neq \tau} \{1 - \cos[s(\psi_{x+\hat{\mu}} - \psi_x)]\} \quad (57)$$

with  $P = \langle \text{Tr } U_\varphi \rangle$ . The sum does not include contributions from links which point in the fixed  $\tau$  direction. If we denote the ‘‘slice’’ of lattice sites with  $\tau$  component  $t$  as  $A_0^\tau = \{x \in A_0 | x_\tau = t\}$ , we see that  $h_1$  is a sum of independent contributions from slices orthogonal to the  $\tau$  direction. Therefore, no smoothness of kernels is needed in the  $\tau$  direction, and from now on we choose  $\psi_x$  to be constant in that direction. Let us assume that the support of  $\psi$  in the  $\tau$  direction is contained in the block  $x'_o$ . Then we find for small  $s$

$$h_1 \approx \frac{1}{2}s^2 \beta P L_B \sum_{x \in A_0^\tau} \sum_{\mu \neq \tau} (\psi_{x+\hat{\mu}} - \psi_x)^2 = \frac{1}{2}s^2 \beta P \alpha_{d-1} \quad (58)$$

with  $\alpha_{d-1} = \langle \psi', -\Delta \psi' \rangle$ . Here,  $\psi'$  denotes the kernel  $\psi$  restricted to the  $(d-1)$ -dimensional sublattice  $A_0^\tau$ , multiplied with a factor  $L_B^{1/2}$  in order to be properly normalized as a  $(d-1)$ -dimensional kernel.

From the kinematical point of view, the behavior of acceptance rates in the  $U(1)$  lattice gauge theory in  $d$  dimensions is the same as in massless free field theory. One might therefore expect that it is possible for a multigrid algorithm to overcome CSD in this model. Indeed, in numerical simulations in two dimensions using piecewise constant kernels, the dynamical critical exponent was found to be  $z \approx 0.1$ .<sup>(13)</sup> However, it was also observed that the multigrid algorithm is not able to move efficiently between different topological sectors. The above-quoted exponent should therefore be interpreted with some care. For the results of a study of a multigrid algorithm for four-dimensional  $U(1)$  theory see ref. 22.

Let us conclude the discussion of the Abelian case with the remark that with no loss of generality one could consider blocks  $x'_o$  that consist only of one layer in the  $\tau$  direction, i.e., effectively  $(d-1)$ -dimensional

blocks. This is a consequence of the fact that the updates of the two variables  $U_{x,\tau}$  and  $U_{y,\tau}$  are statistically independent if  $x_\tau \neq y_\tau$ . This property carries over to the non-Abelian case.

## 7.2. The Non-Abelian Case: Gauge Group $SU(2)$

**7.2.1. Covariant Nonlocal Update Proposal.** We shall now discuss a generalization of the above-described procedure to the non-Abelian case. To be concrete, we study four-dimensional  $SU(2)$  lattice gauge theory.

Let us start with an attempt at a straightforward generalization of the procedure described for the Abelian theory. This would amount to proposing updates

$$U_{x,\tau} \rightarrow U'_{x,\tau} = R_x U_{x,\tau} \tag{59}$$

where the “rotation” matrices  $R_x$  are in  $SU(2)$ . We parametrize them as

$$R_x(\mathbf{n}, s) = \cos(s\psi_x/2) + i \sin(s\psi_x/2) \mathbf{n} \cdot \boldsymbol{\sigma} \tag{60}$$

where  $\mathbf{n}$  denotes a three-dimensional real unit vector, and the  $\sigma_i$  are Pauli matrices.  $\psi$  will have support on three-dimensional blocks  $x'$  of size  $L_B^3$ , and the blocks consist of points lying entirely in  $A_i^\tau$ .

We use the fact that updates of link variables in different slices are statistically independent (as discussed at the end of Section 7.1). One possible updating scheme is to perform the proposed updates on different slices in sequence. Another possible updating scheme consists of building hypercubic four-dimensional blocks out of “staples” of  $L_B$  three-dimensional blocks of size  $L_B^3$  and to perform the updates on this hypercubic block simultaneously. Because of the independence of different slices, the analysis of acceptance rates is the same for both cases. For simplicity we study three-dimensional blocks here.

The energy change associated with the update proposal (59) is

$$\begin{aligned} \Delta\mathcal{H} &= -\frac{\beta}{2} \sum_{\mathcal{D}} \text{Tr}(U'_{\mathcal{D}} - U_{\mathcal{D}}) \\ &= -\frac{\beta}{2} \sum_{x \in A_i^\tau} \sum_{\mu \neq \tau} \text{Tr}[(R_x^* U_{x,\mu} R_{x+\hat{\mu}} - U_{x,\mu}) H_{x,\mu}^*] \end{aligned} \tag{61}$$

with  $H_{x,\mu}^* = U_{x+\hat{\mu},\tau} U_{x+\hat{\tau},\mu}^* U_{x,\tau}^*$ . The relevant quantity for the acceptance rates is  $h_1 = \langle \Delta\mathcal{H} \rangle$ . For piecewise constant kernels  $\psi$  one gets

$$h_1 = \frac{3\beta}{2} A'(L_B - 1) L_B^2 \sin^2(sL_B^{-1/2}/2) + 3\beta PL_B^2 [1 - \cos(sL_B^{-1/2}/2)] \tag{62}$$

with

$$\begin{aligned} A' &= -\langle \text{Tr}[(\mathbf{n} \cdot \boldsymbol{\sigma} U_{x,\mu} \mathbf{n} \cdot \boldsymbol{\sigma} - U_{x,\mu}) H_{x,\mu}^*] \rangle \\ P &= \langle \text{Tr}(U_{x,\mu} H_{x,\mu}^*) \rangle = \langle \text{Tr} U_{\varnothing} \rangle \end{aligned} \quad (63)$$

The first contribution to  $A'$  is the expectation value of a quantity that is not gauge invariant. Determining its gauge-invariant projection, we can show that this contribution vanishes:

$$\langle \text{Tr}(\boldsymbol{\sigma} \cdot \mathbf{n} U_{x,\mu} \boldsymbol{\sigma} \cdot \mathbf{n} H_{x,\mu}^*) \rangle = \int dV \langle \text{Tr}(\mathbf{n} \cdot \boldsymbol{\sigma} V U_{x,\mu} \mathbf{n} \cdot \boldsymbol{\sigma} H_{x,\mu}^* V^*) \rangle = 0 \quad (64)$$

because for  $SU(2)$

$$\int dV \text{Tr}(A V B V^*) = \frac{1}{2} \text{Tr} A \text{Tr} B \quad (65)$$

and the Pauli matrices are traceless. Therefore we get  $A' = P$ .

To the first term in Eq. (62) all links contribute that are entirely inside the block, whereas the second term sums the contributions of all links that have one site in common with the surface of the block. For small  $s$ , the first term behaves like  $s^2 L_B^2$ . This is exactly the behavior of a mass term that, as we learned in the previous sections, is toxic for the multigrid algorithm.

The main difference from the Abelian case is that in addition to the costs from the surface of the block we have a contribution from the interior of the block. Unfortunately, this contribution grows quadratically with the block dimension  $L_B$ .

This does not come as a surprise. Due to the gauge invariance of the model, the  $U$ 's do not have a gauge-invariant meaning. Therefore the rotations  $R_x$  that are smooth over the block for a given gauge can be arbitrarily rough after a gauge transformation. It is therefore clear that the rotation matrices have to be chosen in a gauge-covariant way.

We generalize the update proposal (59) as follows:

$$U_{x,\tau} \rightarrow U'_{x,\tau} = g_x^* R_x g_x U_{x,\tau} \quad (66)$$

with  $g_x \in SU(2)$ . Note that in the Abelian case we obtain nothing new, because  $g_x$  and  $R_x$  commute. In the non-Abelian case, we find for piecewise constant  $\psi$

$$h_1 = \frac{\beta}{2} \sum_{(x, x+\hat{\mu}) \in x'_0} A_{x,\mu} \sin^2(s L_B^{-1/2}/2) + 3\beta P L_B^2 [1 - \cos(s L_B^{-1/2}/2)] \quad (67)$$

with

$$A_{x,\mu} = -\langle \text{Tr}[(\mathbf{n} \cdot \boldsymbol{\sigma} U_{x,\mu}^g \mathbf{n} \cdot \boldsymbol{\sigma} - U_{x,\mu}^g) H_{x,\mu}^{g*}] \rangle \quad (68)$$



Here we have introduced the notation  $U_{x,\mu}^g = g_x U_{x,\mu} g_{x+\hat{\mu}}^*$ . We define  $H^g$  analogously.  $U^g$  is the gauge field obtained by applying a gauge transformation  $g$  to  $U$ . We are free to choose  $g$ . To obtain a valid Monte Carlo algorithm, we require that the  $g$ 's should not depend on the link variables to be updated, i.e., those living on the links  $(x, x + \hat{\tau})$ . On the other hand, we want to minimize  $h_1$ . Let us inspect the quantity  $A_{x,\mu}$  defined in Eq. (68) that leads to the unwanted mass term behavior of  $h_1$ . Consider the extreme case of  $\beta \rightarrow \infty$ . Then the allowed configurations are pure gauges, i.e., configurations that are gauge equivalent to  $U_{x,\mu} = 1$  for all  $x, \mu$ . If we choose  $g$  as the transformation that transforms all links to unity, it is obvious that  $A_{x,\mu}$  is zero. In particular, to obtain this result, it is sufficient to gauge all links inside the block that do not point in the  $\tau$  direction to unity. This consideration leads to following proposal: Choose  $g$  as the gauge transformation that maximizes the functional

$$G_{C,x'_0}(U, g) = \sum_{(x, x + \hat{\mu}) \in x'_0} \text{Tr}(g_x U_{x,\mu} g_{x+\hat{\mu}}^*) \tag{69}$$

We call this gauge the ‘‘block Coulomb gauge.’’ This gauge will bring the links inside the block as close to unity as possible, thus leading to a kind of minimization of  $A_{x,\mu}$  (corresponding to a minimization of the mass term). Note, however, that we do not intend to actually perform the gauge transformation. We use the concept of gauging only to define covariant rotations  $g_x R_x g_x^*$ . Covariance here means that the relevant quantity  $\text{Tr}[(\mathbf{n} \cdot \boldsymbol{\sigma} U_{x,\mu}^g \mathbf{n} \cdot \boldsymbol{\sigma} - U_{x,\mu}^{g*}) H_{x,\mu}^{g*}]$  is now gauge invariant. To see this, assume that we pass from  $U$  to  $U^h$  by applying the gauge transformation  $h$ . The Coulomb gauge condition will then lead to a new  $g' = gh^*$ . Now note that  $(U^h)^{g'h^*} = U^g$ . The same argument applies to  $H$ .

Let us summarize the steps of the nonlocal updating scheme for  $SU(2)$ :

1. Choose a block  $x'_0$  of size  $L_B^3$  that is contained in the slice  $A_t^i$ . All link variables  $U_{x,\tau}$  pointing from sites  $x$  inside the block in the  $\tau$  direction will be moved simultaneously.
2. Find the gauge transformation  $g$  defined by the block Coulomb gauge condition

$$G_{C,x'_0}(U, g) = \sum_{(x, x + \hat{\mu}) \in x'_0} \text{Tr}(g_x U_{x,\mu} g_{x+\hat{\mu}}^*) \stackrel{!}{=} \text{maximal} \tag{70}$$

3. Propose new link variables  $U'_{x,\tau}$  by

$$U_{x,\tau} \rightarrow U'_{x,\tau} = g_x^* R_x g_x U_{x,\tau} \tag{71}$$

with

$$R_x(\mathbf{n}, s) = \cos(s\psi_x/2) + i \sin(s\psi_x/2) \mathbf{n} \cdot \boldsymbol{\sigma} \tag{72}$$

$s$  is a uniformly distributed random number from the interval  $[-\varepsilon, \varepsilon]$ , and  $\mathbf{n}$  is a vector selected randomly from the three-dimensional unit sphere.

4. Calculate the associated change of the Hamiltonian  $\Delta\mathcal{H}$  and accept the proposed link variables with probability  $\min[1, \exp(-\Delta\mathcal{H})]$ .

The detailed balance condition is fulfilled by this updating scheme: For the naive version with  $g=1$  it is straightforward to show that the detailed balance condition holds, since the rotation matrices  $R_x$  are chosen according to a probability distribution which is symmetric around unity.

If we now take  $g$  according to some gauge condition, we have to be careful that we get the same  $g$  before and after the move,  $U_{x,\tau} \rightarrow U'_{x,\tau}$ . Otherwise this move would not be reversible. In other words: The gauge condition yielding  $g$  must not depend on  $U_{x,\tau}$ . This is indeed the case, since only link variables  $U_{x,\mu}$  with  $\mu \neq \tau$  enter in the block Coulomb gauge functional. Note that we do not have to fix the gauge perfectly. If we always use the same procedure in finding  $g$  (e.g., a given number of relaxation sweeps starting from  $g=1$ ), we will always get the same  $g$  and the nonlocal update is reversible.

As usual we now choose different (possibly overlapping) blocks  $x'$ , different block sizes  $L_B$ , different slices  $A_i^\tau$ , and different orientations  $\tau$  of the slices in turn.

### 7.2.2. Acceptance Analysis for Nonlocal $SU(2)$ Updates.

First numerical studies revealed that there is no substantial difference in the acceptance rates when instead of using the block Coulomb gauge condition one uses the Coulomb gauge condition for the whole slice  $A_i^\tau$ :

$$G_C(U, g) = \sum_{(x, x+\bar{\mu}) \in A_i^\tau} \text{Tr}(g_x U_{x,\mu} g_{x+\bar{\mu}}^*) \stackrel{!}{=} \text{maximal} \tag{73}$$

From a practical point of view this gauge condition is very convenient, because the relaxation algorithm to determine the  $g_x$  can then be vectorized in a straightforward way.

If we use the gauge condition (73), the quantity  $A_{x,u}$  becomes translation invariant and also independent of  $\mu$  (where we still keep  $\mu \neq \tau$ ). We get

$$h_1 = (3\beta/2) A(L_B - 1) L_B^2 \sin^2(sL_B^{-1/2}/2) + 3\beta PL_B^2 [1 - \cos(sL_B^{-1/2}/2)] \tag{74}$$

with

$$A = -\langle \text{Tr}[(\mathbf{n} \cdot \boldsymbol{\sigma} U_{x,\mu}^g \mathbf{n} \cdot \boldsymbol{\sigma} - U_{x,\mu}^g) H_{x,\mu}^{g*}] \rangle \tag{75}$$

Following the discussion after Eq. (65), we identify the square root of  $A$  with a “disorder mass”  $m_D$ ,

$$m_D = \sqrt{\beta A} \quad (76)$$

$m_D$  has the dimension of a mass. The presence of a finite  $m_D$  would not be a problem if for large correlation length  $m_D$  scaled like a physical mass (see the discussion for massive free field theory in Section 5).

**7.2.3. Monte Carlo Study of  $m_D$ .** We computed  $m_D$  by Monte Carlo simulations for several values of  $\beta$ . To maximize  $G_C$  we used a standard Gauss–Seidel relaxation algorithm vectorized over a checkerboard structure. The relaxation procedure consists in going through the lattice and minimizing the gauge functional (73) locally. For production runs it would be advantageous to use an accelerated gauge-fixing algorithm such as overrelaxation or multigrid.<sup>(33)</sup> In the Monte Carlo studies reported in this section, we always used 50 Gauss–Seidel sweeps to determine  $g$ . Note that by this procedure,  $G_C$  is not entirely maximized, especially on very large lattices, where the relaxation algorithm suffers from CSD. However, for the detailed balance to be fulfilled, we only need that one uses always the same number of relaxation sweeps. Several tests revealed that increasing the number of relaxation sweeps beyond 50 did not affect the acceptance rates in a substantial way. In our implementation, 50 Gauss–Seidel sweeps over all slices of a given direction  $\tau$  required the same CPU time (on a CRAY Y-MP) as four Creutz heat-bath  $SU(2)$  update sweeps.

We checked the validity of the acceptance formula (18) using Monte Carlo estimates for  $m_D$  and  $P$ . Figure 6 shows results for  $\beta = 2.6$  on a  $20^4$  lattice. The results perfectly justify the usage of the approximation formula. It is therefore sufficient to study the behavior of the quantities  $m_D$  and  $P$ . Our Monte Carlo results are presented in Table III. The last column gives the statistics in sweeps (equilibration sweeps are not counted here). We used a mixture of four microcanonical overrelaxation sweeps followed by a single Creutz heat bath sweep. Measurements (including the determination of  $g$ ) were performed every 25 sweeps.

In Table IV we display the ratios of the disorder mass  $m_D$  with two physical masses, the square root of the string tension  $\kappa$  and the lowest glue ball mass  $m_{0+}$ . The estimates for the physical masses are taken from ref. 34. The results show that the disorder mass is nearly independent of  $\beta$  in the range studied, whereas the physical masses decrease by roughly a factor of two. Thus,  $m_D$  is not scaling like a physical mass for the couplings studied here. We conclude from this that for large blocks the term quadratic in

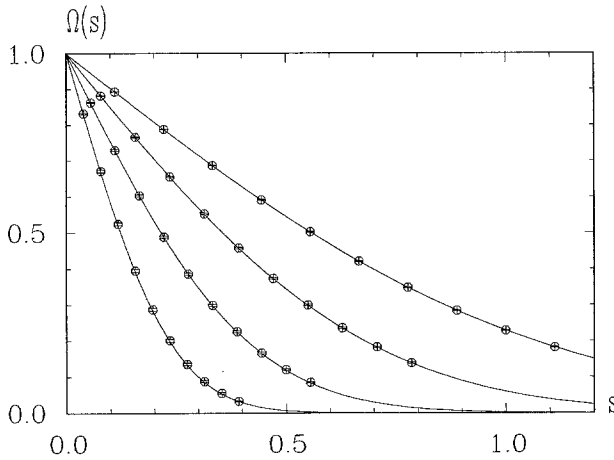


Fig. 6.  $\Omega(s)$  in four-dimensional  $SU(2)$  lattice gauge theory using piecewise constant kernels,  $\beta = 2.6$  on a  $20^4$  lattice. From top to bottom:  $L_B = 2, 4, 8,$  and  $16$ . Points with error bars: Monte Carlo results; lines: analytical results using  $m_D$  and  $P$  from Monte Carlo (errors smaller than line width).

$L_B$  will strongly suppress the acceptance rates even when the ratio of correlation length and block size  $L_B$  is kept constant.

However, one could hope that the value of the unwanted mass term is so small that it does no harm in practical calculations. Let us examine the effect of this mass term in more detail. Recall that  $h_1$  is built up from two contributions. The first contribution is that related to the gauge field disorder inside the block and is quantitatively represented by the mass  $m_D$ . The second contribution is associated with the block surface. The latter can of course be made smaller by using smooth kernels  $\psi$  instead of the piecewise constant kernels discussed so far. However, the disorder term

Table III. Monte Carlo Results for  $m_D$  and  $P$

Lattice size	$\beta$	$m_D$	$P$	Statistics
$8^4$	2.4	0.717(3)	1.2609(5)	10,000
$12^4$	2.4	0.7010(6)	1.2601(3)	10,000
$16^4$	2.4	0.7007(3)	1.2599(1)	10,000
$8^4$	2.6	0.703(6)	1.3407(3)	30,000
$12^4$	2.6	0.657(2)	1.3403(2)	20,000
$16^4$	2.6	0.6568(5)	1.3401(1)	10,000
$20^4$	2.6	0.6576(3)	1.3402(1)	5,000

Table IV. Comparison of  $m_D$  with Physical Masses

Lattice size	$\beta$	$m_D$	$\sqrt{\kappa}$	$m_{0+}$	$m_D/\sqrt{\kappa}$	$m_D/m_{0+}$
$16^4$	2.4	0.7007(3)	0.258(2)	0.94(3)	2.72	0.75
$20^4$	2.6	0.6576(3)	0.125(4)	0.52(3)	5.26	1.26

cannot be expected to become smaller for smooth kernels (see below). In Fig. 7 we plot separately the two contributions to  $h_1$ ,

$$\begin{aligned}
 h_{1,A} &= (3\beta/2) A(L_B - 1) L_B^2 \sin^2(sL_B^{-1/2}/2) \\
 h_{1,P} &= 3\beta PL_B^2 [1 - \cos(sL_B^{-1/2}/2)]
 \end{aligned}
 \tag{77}$$

for  $\beta=2.6$  and block size  $L_B=8$  on a  $20^4$  lattice. The plot shows that already for this block size the disorder contribution is by no means small—it is comparable to the surface effect. It is therefore not clear that one could achieve any significant improvement by using smooth kernels. To investigate this in more detail, we derive an expression for  $h_1$ , valid for smooth kernels as well:

$$h_1 = \frac{3\beta}{8} s^2 A \sum_{x \in A_i^+} \psi_x^2 + \frac{\beta}{16} s^2 (P - A) \alpha_3 + O(s^4)
 \tag{78}$$

Since  $\sum \psi_x^2 \sim L_B^2$ , we get essentially the same behavior for the disorder contribution as in the case of piecewise constant kernels.

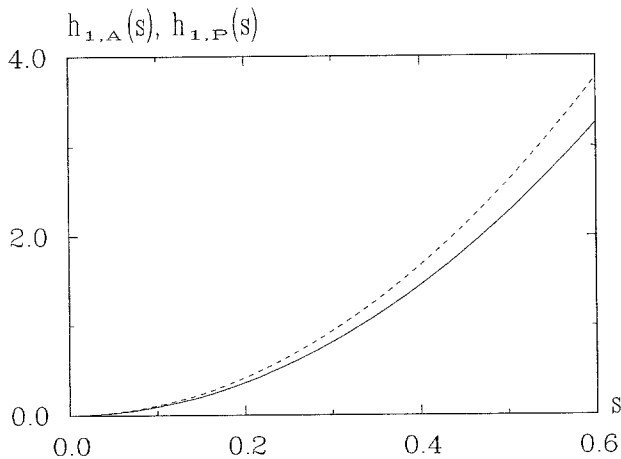


Fig. 7. Comparison of disorder and surface effects for four-dimensional  $SU(2)$  lattice gauge theory using piecewise constant kernels on an  $8^3$  block,  $\beta=2.6$  on a  $20^4$  lattice. Solid line:  $h_{1,A}(s)$  (disorder effects); dashed line:  $h_{1,P}(s)$  (surface effects).

For smooth  $\psi^{\text{sine}}$  kernels we show separately in Fig. 8 the two contributions to  $h_1$ ,

$$h_{1,A} = \frac{3\beta}{8} s^2 A \sum_{x \in A_i^t} \psi_x^2, \quad h_{1,P-A} = \frac{\beta}{16} s^2 (P-A) \alpha_3 \quad (79)$$

for  $\beta = 2.6$  and block size  $L_B = 8$  on a  $20^4$  lattice. We observe that the surface effects are lowered by the smooth kernels, but the disorder contribution is even higher than for piecewise constant kernels.

Piecewise constant kernels have the practical feature that once the change of the Hamiltonian has been calculated, one can perform multihit Metropolis updating or microcanonical overrelaxation. In a special case, even an explicit multigrid implementation with a W-cycle is possible (see below). For smooth kernels the change in the Hamiltonian would have to be calculated again and again. Also, the advantage of smooth kernels is not that clear on small three- or four-dimensional blocks. For an actual simulation we would therefore prefer piecewise constant kernels.

**7.2.4. Maximally Abelian Gauge.** Our proposal for the choice of  $g$  was motivated by the desire to minimize the quantity  $A$ . We now ask whether there is a better choice of  $g$  than the  $g$  determined by the Coulomb gauge condition. For the sake of simplicity let us take  $\mathbf{n} = (0, 0, 1)$ , i.e.,  $\mathbf{n} \cdot \boldsymbol{\sigma} = \sigma_3$ . Then  $A$  is given by

$$A = -\langle \text{Tr}[(\sigma_3 U_{x,\mu}^g \sigma_3 - U_{x,\mu}^g) H_{x,\mu}^{g*}] \rangle \quad (80)$$

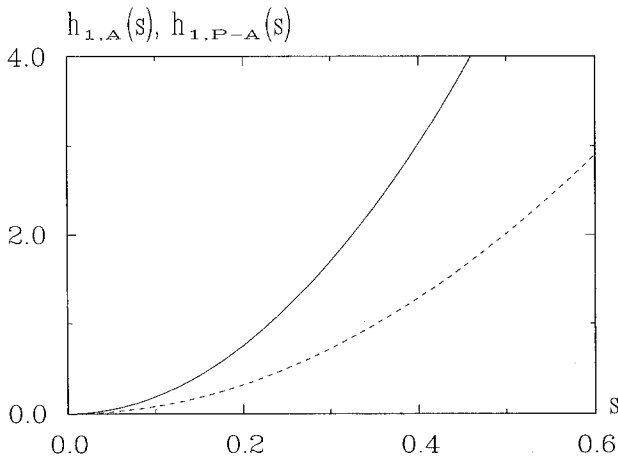


Fig. 8. Comparison of disorder and surface effects for the four-dimensional  $SU(2)$  lattice gauge theory using smooth  $\psi^{\text{sine}}$  kernels on an  $8^3$  block,  $\beta = 2.6$  on a  $20^4$  lattice, quadratic approximation used. Solid line:  $h_{1,A}(s)$  (disorder effects); dashed line:  $h_{1,P-A}(s)$  (surface effects).

The choice of the Coulomb gauge condition aimed at bringing  $U_{x,\mu}^g$  as close to unity as possible. Alternatively, one might require that  $U_{x,\mu}^g$  should be as close as possible to an  $SU(2)$  matrix of the form  $a_0 + ia_3\sigma_3$ . This will also lead to a small  $A$ . The corresponding gauge transformation  $g$  can be found by maximizing the functional

$$G_A(U, g) = \sum_{(x, x+\hat{\mu}) \in A_1^i} \text{Tr}(\sigma_3 U_{x,\mu}^g \sigma_3 U_{x,\mu}^{g*}) \quad (81)$$

leading to the maximally Abelian gauge,<sup>(35)</sup> here implemented only on a slice. We computed  $m_D$  also using the  $g$ 's resulting from this gauge condition and compared the results with the ones obtained by using the Coulomb gauge condition. We did not find a substantial difference. We prefer the Coulomb gauge condition because it does not depend on the direction  $\mathbf{n}$  and thus saves computer time.

**7.2.5. Proposal for an Implementation.** An explicit multigrid implementation is possible in a special case if we use piecewise constant kernels. This was pointed out in ref. 22 in a related context.

The idea is to update only a fixed  $U(1)$  subgroup of  $SU(2)$  globally: We divide the fundamental lattice  $A_0$  into hypercubic blocks  $x'$  of size  $2^4$  and “rotate” all links going from sites  $x \in x'$  in a fixed  $\tau$  direction with the same angle  $\theta_{x'}$ :

$$U_{x,\tau} \rightarrow U'_{x,\tau} = g_x^* R_x g_x U_{x,\tau} \quad (82)$$

with

$$R_x(\mathbf{n}_{x'}, \theta_{x'}) = \cos(\theta_{x'}/2) + i \sin(\theta_{x'}/2) \mathbf{n}_{x'} \cdot \boldsymbol{\sigma} \quad (83)$$

The gauge transformation  $g$  is obtained by imposing the Coulomb gauge condition on slices  $A_1^i$  as defined above. We now consider the special case where the directions  $\mathbf{n}_{x'}$  of the embedded  $U(1)$  subgroups are independent of the block  $x'$ , i.e.,  $\mathbf{n}_{x'} = \mathbf{n}$  for all  $x' \in A_1$ . Then we get a conditional Hamiltonian  $\mathcal{H}_1(\theta)$  by substituting the “rotated” gauge field  $U'$  in the fundamental Hamiltonian. By iterating this procedure, one gets conditional Hamiltonians  $\mathcal{H}_k(\theta)$  on coarser layers  $A_k$ . The point is that in the special case considered here  $\mathcal{H}_k(\theta)$  always stays of the form

$$\begin{aligned} -\mathcal{H}_k(\theta) = & \frac{1}{2} \sum_{x' \in A_k} \{ \beta_{x'}^{cc} \cos^2(\theta_{x'}) + \beta_{x'}^{cs} \cos(\theta_{x'}) \sin(\theta_{x'}) + \beta_{x'}^{ss} \sin^2(\theta_{x'}) \} \\ & + \frac{1}{2} \sum_{x' \in A_k} \sum_{\mu \neq \tau} \{ \beta_{x',\mu}^{cc} \cos(\theta_{x'}) \cos(\theta_{x'+\hat{\mu}}) + \beta_{x',\mu}^{cs} \cos(\theta_{x'}) \sin(\theta_{x'+\hat{\mu}}) \\ & + \beta_{x',\mu}^{sc} \sin(\theta_{x'}) \cos(\theta_{x'+\hat{\mu}}) + \beta_{x',\mu}^{ss} \sin(\theta_{x'}) \sin(\theta_{x'+\hat{\mu}}) \} \\ & + \text{const} \end{aligned} \quad (84)$$

with space-dependent couplings that can be recursively calculated from the couplings defined on the next finer layer  $A_{k-1}$ . Therefore, a W-cycle is possible. Of course, this implementation is also possible with three-dimensional blocks.

## 8. SUMMARY

We have presented a simple yet accurate formula that expresses acceptance rates for nonlocal update algorithms in terms of one single parameter (or two in the case of non-Abelian gauge theory) entering the quantity  $h_1$ . This parameter is easy to compute, e.g., by Monte Carlo simulations on a small lattice. We encountered two classes of models. For sine-Gordon,  $\phi^4$  theory, and  $SU(2)$  lattice gauge theory,  $s$  had to be rescaled like  $L_B^{-1}$  for piecewise constant and for smooth kernels, whereas for massless free field theory, the  $XY$  model, the  $O(N)$  nonlinear  $\sigma$ -model, and  $U(1)$  lattice gauge theory one can achieve  $L_B$ -independent acceptance rates by choosing smooth kernels.

We can compare the behavior of the acceptance rates in interacting models with free field theory, where CSD is known to be eliminated by a multigrid algorithm. In order to do this, we presented a study of the influence of the coarse-to-fine interpolation on the acceptance rates in free field theory.

The results of the comparison are consistent with the following rule: For an interacting model, sufficiently high acceptance rates for a complete elimination of CSD can only be expected if  $h_1 = \langle \mathcal{H}(\phi + s\psi) - \mathcal{H}(\phi) \rangle$  contains no algorithmic “mass” term  $\sim s^2 \sum_x \psi_x^2$ . With the help of this rule it is possible to decide whether a given statistical model is a natural candidate for multigrid Monte Carlo or not.

The kinematical mechanism that can lead to a failure of direct generalizations of multigrid algorithms from free field theory to interacting models is well described by our analysis. We hope that a better understanding can lead to improved multigrid algorithms that can overcome kinematical obstructions stemming from an algorithmic “mass” term.

The acceptance rates of our proposal for nonlocal updates in  $SU(2)$  lattice gauge theory were investigated in detail. Here we found that an algorithmic “mass” term generated by the disorder in the gauge field suppresses the acceptance rates on large blocks. From this study we do not expect that our algorithm will have a chance to overcome CSD. We think that the best method for a numerical experiment in order to check our prediction would be an explicit multigrid implementation using piecewise constant kernels and a W-cycle. An implementation and test is planned.



The crucial question is, of course, whether one is able to beat the over-relaxation algorithm.

## ACKNOWLEDGMENTS

Our study on multigrid Monte Carlo algorithms for non-Abelian gauge theory was inspired by related work of Thomas Kalkreuter, Gerhard Mack, and Steffen Meyer. We would like to thank them for many stimulating discussions. The idea of our  $SU(2)$  algorithm was developed together with Martin Hasenbusch. It is a pleasure for us to thank him for the enjoyable collaboration.

Our  $SU(2)$  program contains some  $SU(2)$  update routines written by Hans Gerd Evertz and Mihai Marcu. The numerical computations were performed on the NEC SX-3 in Cologne, the Landesvektorrechner of the RWTH in Aachen, and the CRAY Y-MP of the HLRZ in Jülich. One of us (M.G.) would like to thank the Deutsche Forschungsgemeinschaft for financial support.

## NOTE ADDED IN PROOF

We implemented and tested our proposal for  $SU(2)$  gauge fields in four dimensions, using an explicit multigrid algorithm with three-dimensional blocks, as described in Section 7.2.5. On an  $8^4$  lattice, with  $\beta=2.4$  and  $\beta=2.6$ , we did not observe any substantial speed up compared to a local heat bath algorithm (not even a constant factor).

## REFERENCES

1. K. Binder, ed., *Monte Carlo Methods in Statistical Physics* (Springer, Berlin, 1986).
2. C. Rebbi, ed., *Lattice Gauge Theory and Monte Carlo Simulations* (World Scientific, Singapore, 1983).
3. S. L. Adler, *Phys. Rev. D* **23**:458 (1988); H. Neuberger, *Phys. Rev. Lett.* **59**:1877 (1987).
4. U. Wolff, in *Proceedings of the 31. IUKT, Schladming, Austria, February 1992*, H. Gausterer and C. B. Lang, eds. (Springer, Berlin, 1992).
5. A. D. Kennedy and B. Pendleton, in *Lattice '90* (Proceedings of the International Symposium, Tallahassee, Florida, 1990), U. M. Heller, A. D. Kennedy, and S. Sanielevici, eds., *Nucl. Phys. B (Proc. Suppl.)* **20**:118 (1991); S. Gupta, in *Lattice '91* (Proceedings of the International Symposium, Tsukuba, Japan), M. Fukugita *et al.*, eds., *Nucl. Phys. B (Proc. Suppl.)* **26**:617 (1992).
6. R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **58**:86 (1987); U. Wolff, *Phys. Rev. Lett.* **62**:361 (1989).
7. G. Parisi, in *Progress in Gauge Field Theory* (Proceedings of the Cargèse Summer Institute, Cargèse, France, 1983), G. 't Hooft *et al.*, eds. (Plenum Press, New York, 1984).
8. J. Goodman and A. D. Sokal, *Phys. Rev. Lett.* **56**:1015 (1986).

9. G. Mack, in *Nonperturbative Quantum Field Theory* (Proceedings, Cargèse, France, 1987), G. 't Hooft *et al.*, eds. (Plenum Press, New York, 1988).
10. U. Wolff, in *Workshop on Fermion Algorithms* (HLRZ, KFA Jülich, April 1992), H. J. Herrmann and F. Karsch, eds. (World Scientific, Singapore, 1991).
11. M. Hasenbusch and S. Meyer, *Phys. Rev. Lett.* **68**:435 (1992).
12. M. Hasenbusch, G. Mack, and S. Meyer, in Lattice '90 (Proceedings of the International Symposium, Tallahassee, Florida, 1990), U. M. Heller, A. D. Kennedy, and S. Sanielevici, eds., *Nucl. Phys. B (Proc. Suppl.)* **20**:110 (1991).
13. M. L. Laursen, J. Smit, and J. C. Vink, *Phys. Lett. B* **262**:467 (1991).
14. R. G. Edwards, S. J. Ferreira, J. Goodman, and A. D. Sokal, *Nucl. Phys. B* **380**:621 (1992).
15. J. Goodman and A. D. Sokal, *Phys. Rev. D* **40**:2035 (1989).
16. J. Linn, Diploma thesis, Kaiserslautern (1991).
17. M. Grabenstein and K. Pinn, *Phys. Rev. D* **45**:4372 (1992).
18. M. Creutz, *Phys. Rev. D* **36**:515 (1987); F. R. Brown and T. J. Woch, *Phys. Rev. Lett.* **58**:2394 (1987).
19. S. L. Adler, in Lattice '88 (Proceedings of the International Symposium, Batavia, Illinois, 1988), A. S. Kronfeld and P. B. Mackenzie, eds., *Nucl. Phys. B (Proc. Suppl.)* **9**:437 (1989).
20. R. Ben-Av, H. G. Evertz, M. Marcu, and S. Solomon, *Phys. Rev. D* **44**:2953 (1991).
21. R. Sinclair, *Phys. Rev. D* **45**:2098 (1992).
22. M. L. Laursen and J. C. Vink, preprint HLRZ-92-14 (March 1992).
23. G. Bhanot and S. L. Adler, *Phys. Rev. Lett.* **66**:1806 (1991); G. Bhanot, S. L. Adler, T. Lippert, K. Schilling, and P. Ueberholz, *Nucl. Phys. B* **368**:745 (1992).
24. A. Brandt, in Lattice '91 (Proceedings of the International Symposium, Tsukuba, Japan), M. Fukugita *et al.*, eds., *Nucl. Phys. B (Proc. Suppl.)* **26**:137 (1992).
25. K. Stüben and U. Trottenberg, in *Multigrid Methods* (Proceedings, Köln-Porz, 1981), W. Hackbusch *et al.*, eds. (Springer, Berlin, 1982).
26. S. Gupta, A. Irbäck, F. Karsch, and B. Petersson, *Phys. Lett. B* **242**:437 (1990).
27. K. Gawędzki and A. Kupiainen, *Commun. Math. Phys.* **77**:31 (1980); **99**:197 (1985), and references cited therein.
28. G. Mack and S. Meyer, in Lattice '89 (Proceedings of the International Symposium, Capri, Italy, 1989), R. Petronzio *et al.*, eds., *Nucl. Phys. B (Proc. Suppl.)* **17**:293 (1990).
29. T. Kalkreuter, *Nucl. Phys. B* **376**:637 (1992).
30. G. Mack, T. Kalkreuter, G. Palma, and M. Speh, in *Proceedings of the 31. IUKT, Schladming, Austria, February 1992*, H. Gausterer and C. B. Lang, eds. (Springer, Berlin, 1992).
31. M. Hasenbusch, M. Marcu, and K. Pinn, in Lattice '91 (Proceedings of the International Symposium, Tsukuba, Japan), M. Fukugita *et al.*, eds., *Nucl. Phys. B (Proc. Suppl.)* **26**:598 (1992).
32. R. G. Edwards, J. Goodman, and A. D. Sokal, *Nucl. Phys. B* **354**:289 (1991); A. Hulsebos, J. Smit, and J. C. Vink, *Nucl. Phys. B* **356**:775 (1991).
33. J. E. Mandula and M. Ogilvie, *Phys. Lett. B* **248**:156 (1990); A. Hulsebos, M. L. Laursen, J. Smit, and A. J. van der Sijs, in Lattice '90 (Proceedings of the International Symposium, Tallahassee, Florida, 1990), U. M. Heller, A. D. Kennedy, and S. Sanielevici, eds., *Nucl. Phys. B (Proc. Suppl.)* **20**:110 (1991).
34. C. Michael and M. Teper, *Phys. Lett. B* **199**:95 (1987).
35. A. S. Kronfeld, M. L. Laursen, G. Schierholz, and U.-J. Wiese, *Phys. Lett.* **198B**:516 (1987).