

## Optimal Multigrid Algorithms for Variable-Coupling Isotropic Gaussian Models

A. Brandt<sup>1</sup> and M. Galun<sup>1</sup>

Received July 31, 1996; final February 19, 1997

---

A novel class of multigrid algorithms for the variable-coupling isotropic Gaussian models is presented. In addition to the elimination of the critical slowing down (which otherwise might become much worse than usual in the case of strongly varying coupling values), the "volume factor" is also eliminated. That is, the need to produce many independent fine-grid configurations for averaging out their statistical deviations is removed, by applying multigrid cycles that sample mostly on coarse grids. Thermodynamic limits can be calculated to relative accuracy  $\varepsilon_r$  in just  $O(\varepsilon_r^{-2})$  computer operations, where  $\varepsilon_r$  is the error relative to the standard deviation of the observable. In this paper, such an optimal algorithm is obtained for the calculation of the susceptibility in the  $d$ -dimensional variable-coupling isotropic Gaussian model (with numerical experiments for  $d = 1, 2$ ). Some basic general rules for the operation of multigrid algorithms, applicable to much wider classes of models, are derived.

---

**KEY WORDS:** Multigrid; Monte Carlo; critical slowing down; volume factor; thermodynamic limit; variable-coupling isotropic Gaussian model.

### 1. INTRODUCTION

One of the aims in statistical physics is to calculate various average properties of configurations governed by the Boltzmann distribution. This is usually done by measuring these averages over a sequence of Monte Carlo iterations. Unfortunately, such processes tend to suffer from *several* independent inefficiency factors that multiply each other and thus produce very expensive computations.

The best known of these inefficiency factors is the *critical slowing down* (CSD). This is the phenomenon, typical to critical systems, that with the

---

<sup>1</sup> Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel; e-mail: meirav@wisdom.weizmann.ac.il.

increase in lattice size there also comes an increase in the number of full Monte Carlo passes over the lattice needed to produce a new configuration which is statistically “useful,” i.e., substantially independent of, or only weakly correlated to, a former configuration. More precisely, the process requires  $O(N^z)$  Monte Carlo sweeps, hence  $O(N^{d+z})$  computer operations, to create a new independent configuration, where  $N$  is the linear lattice size,  $d$  is the dimension and  $z > 0$  is the CSD exponent (typically  $z \approx 2$ ). Considerable efforts have been devoted to reduce the critical slowing down. For simple cases with real variables, classical multigrid methods<sup>(12, 17, 21)</sup> can eliminate the CSD (i.e., obtain  $z = 0$ ). For more complicated models, (e.g.,  $\phi^4$ , nonlinear  $\sigma$ -models or discrete models) more recent publications report on simulation techniques that partially<sup>(14, 15, 17, 23, 24, 26)</sup> or completely<sup>(5, 18–20, 22, 27)</sup> eliminate the CSD. This means that the computer work to produce an independent configuration is proportional to the number of gridpoints, i.e.,  $O(N^d)$  operations.

This paper treats the Gaussian models with non-constant couplings, therefore of special interest are cases where the couplings change strongly from one subdomain to another. In such cases, the usual critical slowing down of the point-by-point Monte Carlo process is compounded by a very severe sampling slowness, i.e., the number of sweeps for producing an independent configuration may grow as  $O(a_* N^2)$  where  $a_*$  is the maximal ratio between the values of the coupling.

In addition to the CSD factor  $N^z$  there is another, no less important factor of slowness: namely, the above  $N^d$  factor, called the *volume factor*. Indeed, to calculate a thermodynamic quantity to a certain *relative* accuracy  $\varepsilon_r$ , one needs to produce  $O(\varepsilon_r^{-2})$  essentially independent configurations to average out the deviation exhibited by each of them, where the *relative* accuracy  $\varepsilon_r$  is the error relative to  $\sigma$ , the standard deviation of the observable in question. Also, the size  $N^d$  of the grid must increase as some positive power of  $\varepsilon_r^{-1}$ . Thus, even if the CSD has been completely eliminated, the overall work increases as  $O(\varepsilon_r^{-2} N^d)$ . An important advantage of the multigrid approach is that it can drastically reduce the volume factor  $N^d$  as well, by averaging over many samples produced on coarse levels of the multigrid cycle. Actually, even for extreme cases of large  $a_*$  we will demonstrate below that by suitable cycling and sampling procedures one can completely remove both the volume factor and the compounded CSD mentioned above.

The elimination of both the volume factor and the CSD factor means that a thermodynamic limit can be calculated to an accuracy of  $\pm \varepsilon$  in *optimal time*, i.e., in only  $O(\sigma^2 \varepsilon^{-2})$  computer operations. This is just the same order of complexity as needed to calculate, by statistical trials, any simple “pointwise” average, such as the frequency of “heads” in coin

tossing. By contrast, both the volume and the CSD factors *multiply* the statistical factor ( $\sigma^2 \varepsilon^{-2}$ ) in the operation count of conventional algorithms

The elimination of the volume factor has first been demonstrated<sup>(7,10,16)</sup> for the Gaussian model with constant coefficients. It has been shown there, for the one-dimensional Gaussian model, that the susceptibility can be calculated to accuracy  $\varepsilon_r$  in about  $4\varepsilon_r^{-2}$  random number generations, while the average energy per degree of freedom requires  $3\varepsilon_r^{-2}$  such generations for a similar accuracy. In the two-dimensional Gaussian model, the susceptibility can be measured to accuracy  $\varepsilon_r$  in about  $20\varepsilon_r^{-2}$  random number generations. Moreover, we have shown for the one dimensional *massive* Gaussian model<sup>(9)</sup> that the susceptibility is calculated to relative accuracy  $\varepsilon_r$  in less than  $8\varepsilon_r^{-2}$  random generations, essentially independently of the mass size, although the algorithm flow does change with that size.

These previous calculations have not provided convincing demonstration of the power and generality of the approach, because the constant-coefficient Gaussian models in rectangular (or periodic, domains can be treated with similar efficiency also by Fourier-based algorithms or by closed-form Fourier analysis (as indeed used in the aforementioned works<sup>(7,9,10,16)</sup>) both for calculating the desired physical quantities and for analyzing the multigrid algorithms). To demonstrate more general applicability of the multigrid approach, the elimination of both the CSD and the volume factor is shown here for variable-coefficient cases for which Fourier methods are inapplicable.

A multigrid simulator for the variable-coupling Gaussian model provides an important basis for general nonlinear models, where non-constant couplings stochastically emerge at coarser levels of the multigrid Monte Carlo processing. Indeed, in a companion paper,<sup>(8)</sup> the removal of the CSD and the volume factor is shown for some cases of a simple *non-linear* model—the anharmonic crystal.

We show that in order to reach optimality in the variable-coupling Gaussian model, the multigrid algorithm for such cases must differ from the algorithms in refs. 7, 9, 10, and 16 mainly in the following two aspects. Firstly, instead of the simple linear interpolation that we have used in refs. 7, 9, 10, and 16, *weighted interpolation* must be used. Secondly, *variable sampling* should be applied during the multigrid cycle; in particular, the Monte Carlo process should sample more frequently regions with smaller coupling values. Precise rules for the interpolation weights and for the sampling frequency (implying also general rules for switching between the multigrid levels) are derived below.

The algorithm have been implemented for strongly discontinuous cases (large  $a_*$ ) in one and two dimensions. The results are as good as those previously attained<sup>(7,9,10,16)</sup> for constant coefficients. For the one

dimensional variable-coupling Gaussian model, the susceptibility is calculated to accuracy  $\varepsilon_r$  in about less than  $8\varepsilon_r^{-2}$  random number generations. In the two-dimensional variable-coupling Gaussian model, the susceptibility can be measured in about less than  $20\varepsilon_r^{-2}$  random generations. These results are independent of the coupling ratio  $a_*$ .

Thus, our multigrid algorithm effectively produces an independent sample in just  $O(1)$  computer operations. The computational time of this “optimal multigrid” algorithm is thus smaller by a factor  $O(N^d)$  compared to that of a *conventional* multigrid algorithm, e.g., such as that of refs. 12, 17, and 21, which measures the observable only once per multigrid cycle.

For simplicity, the present work deals only with *isotropic* models. Modifications to the anisotropic case are briefly discussed in Appendix B.

## 2. VARIABLE-COUPLING ISOTROPIC GAUSSIAN MODEL IN GENERAL DIMENSION

### 2.1. Continuum and Discrete Models

The general variable-coupling  $d$ -dimensional isotropic Gaussian Hamiltonian is defined in the continuum by

$$\mathcal{H}(u) = \int_{\Omega} a(x) \left( \left( \frac{\partial u}{\partial x^1} \right)^2 + \dots + \left( \frac{\partial u}{\partial x^d} \right)^2 \right) dx^1 \dots dx^d \quad (1)$$

where  $u = u(x) = u(x^1, \dots, x^d)$  and  $a(x)$  are real functions defined for  $x = (x^1, \dots, x^d) \in \Omega$ ;  $\Omega$  is a domain in  $R^d$  on the boundary of which values of  $u$  are prescribed. For definiteness, we assume homogeneous Dirichlet ( $u = 0$ ) boundary conditions.

The magnetization is defined here as

$$M(u) = \frac{1}{|\Omega|} \int_{\Omega} u(x) dx^1 \dots dx^d \quad (2)$$

where  $|\Omega|$  is the volume of  $\Omega$ , and the probability density of the configurations is the Boltzmann distribution

$$P(u) = \frac{1}{Z} e^{-\mathcal{H}(u)} \quad (3)$$

where the temperature  $T$  is absorbed in  $\mathcal{H}(u)$  and  $Z$  is a normalization factor (the partition function) derived from the condition  $\int_u P(u) du = 1$ . As

the density function is given, average properties of interest are the average magnetization  $\langle M \rangle = \int_{\Omega} M(u) P(u) du$  and the susceptibility

$$\chi = \langle M^2 \rangle - \langle M \rangle^2 \quad (4)$$

Clearly, in the case of the sign-symmetric Hamiltonian (1) and the homogeneous boundary conditions,  $\langle M \rangle = 0$ .

Discrete approximations will be calculated by placing a grid  $\Omega^h$  of points  $x_i$  over the domain  $\Omega$ , where  $i = (i^1, \dots, i^d)$  is a vector of integers,  $x_i = (x_i^1, \dots, x_i^d) = (i^1 h, \dots, i^d h)$ ,  $h$  being the (real and positive) meshsize. The value of the discrete configuration  $u^h$  at the point  $x_i$  will be denoted  $u_i^h$ . For simplicity, we will assume that the boundary of  $\Omega$  is a union of pieces each of which is included in a grid hyperplane, so that boundary conditions can be discretized in the most direct and obvious way; e.g.,  $u_i^h = 0$  for  $x_i$  in the boundary; but extensions to more general cases are quite straightforward. The discrete Hamiltonian and magnetization will be given, respectively, by the second-order discretizations

$$\mathcal{H}_h(u^h) = h^{d-2} \sum_{\langle i, j \rangle} a_{ij} (u_i^h - u_j^h)^2 \quad (5)$$

and

$$M_h(u^h) = \frac{h^d}{|\Omega|} \sum_i u_i \quad (6)$$

where  $\langle i, j \rangle$  is any pair of nearest-neighbor sites, including the case that one of them is on the boundary, and  $\sum_i$  runs over all interior sites. (In case of non-homogeneous boundary conditions, to obtain a second order approximation, boundary values multiplied by 1/2 should be added to the sum in (6).) The coupling coefficient  $a_{ij}$  is a proper homogenization of  $a(x)$ ; i.e., it represents harmonic averaging (the inverse of the average of  $a(x)^{-1}$ ) in the  $i$  to  $j$  direction, compounded with usual averaging in the perpendicular directions. Similarly to the continuous case, the probability distribution is given by (3), with  $\mathcal{H}_h(u^h)$  replacing  $\mathcal{H}(u)$ . The discrete averages which estimates the continuum averages are naturally the discrete average magnetization  $\langle M_h \rangle$  and the discrete susceptibility

$$\chi_h = \langle M_h^2 \rangle - \langle M_h \rangle^2 = \langle M_h^2 \rangle \quad (7)$$

Fourier expansions were used for the constant-coefficient Gaussian model<sup>(7, 10, 16)</sup> ( $a(x) \equiv 1$ ) and the massive Gaussian model<sup>(9)</sup> to compute continuum and discrete averages analytically and to construct an optimal

multigrid algorithm. When  $a(x)$  is not constant, or when  $\Omega$  is not rectangular, Fourier expansions can no longer serve, neither for exact calculations of continuum (thermodynamic limit) and discrete averages, nor for analyzing the multigrid Monte Carlo simulations. A more general way to analyze the multigrid Monte Carlo algorithm in the variable-coupling Gaussian models is described below (see Sec. 2.5).

## 2.2. Extreme Monte Carlo Slowness

In some cases the usual critical slowing down of the point-by-point Monte Carlo simulation is compounded by a very severe sampling slowness resulting from widely differing values of the coupling  $a_{ij}$ . For example, if  $a(x) = 1$  in a region all around the boundary while  $a(x) = a_*$  in some interior subdomain, then when  $a_*$  is getting large the interior spin block becomes strongly coupled. Therefore, point-by-point relaxation would allow only small fluctuations in this spin block, although a uniform movement of the block is physically probable. Clearly, as  $a_*$  increases a uniform block movement becomes much slower. Generally, for this reason, the decorrelation time may increase proportionately to  $a_*$ ; i.e., the number of Monte Carlo sweeps required to produce a new effectively independent configuration may increase like  $O(a_* N^2)$ , where  $N = O(h^{-1})$  is the typical number of grid points in each coordinate. Thus, for large  $a_*$  there is additional reason for constructing an accelerated multigrid algorithm.

We demonstrate this compounded slowdown of point-by-point simulations by an example. We have measured the simulation efficiency by the integrated autocorrelation time

$$\tau = \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho(t) \quad (8a)$$

as determined from the normalized autocorrelation function

$$\rho(t) = \frac{\langle A_i A_{i+t} \rangle_i - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2} \quad (8b)$$

where  $A_i$  is the measurement of  $M^2$  produced by the simulation at the end of the  $i$ th sweep over the entire lattice. As an example we consider the one-dimensional variable-coupling Gaussian model in the interval  $[0, 1]$  with the step function  $a(x)$  determined by:

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.25 \\ a_* & \text{if } 0.25 \leq x \leq 0.75 \\ 1 & \text{if } 0.75 < x \leq 1 \end{cases} \quad (9)$$

**Table I. Measuring the Autocorrelation Time  $\tau$  for the Susceptibility in the Monte Carlo Process**

$N$	$a_* = 1$	$a_* = 10$	$a_* = 100$	$a_* = 1000$
4	1	5	48	515
8	3	20	208	18200
16	12	75	784	
32	54	250		

We have measured  $\tau$  for the point-by-point Monte Carlo process for different values of  $a_*$ . Results are presented in Table I, showing that for  $a_* \gg 1$ , the autocorrelation time  $\tau$  is indeed proportional to  $a_* N^2$ , where  $N = 1/h$ .

### 2.3. Description of the Multigrid Cycle

Consider the one-dimensional variable-coupling Gaussian model in the interval  $[0, L]$ . The generalized discretized Hamiltonian,  $\mathcal{H}_h(u^h)$ , on a grid with meshsize  $h = L/N$ , can be written as

$$\mathcal{H}_h(u) = \frac{1}{h} \sum_{i=1}^N a_i (u_i - u_{i-1})^2 + h \sum_{i=1}^{N-1} \phi_i u_i \tag{10}$$

where  $u_i$  is the variable at the gridpoint  $x_i = ih$ ,  $0 \leq i \leq N$ , and  $a_i$  denotes the coupling between  $u_{i-1}$  and  $u_i$ ,  $1 \leq i \leq N$ . For simplicity of the multigrid algorithm we assume  $N = 2^k$ . On the finest grid,  $\phi_i = 0$  ( $i = 1, \dots, N - 1$ ), but the more general form of the Hamiltonian is needed for the algorithm recursion.

The coarse grid with meshsize  $H = 2h$  is constructed by taking every other grid-point. The coarse-grid function  $u^H = (u_0^H, \dots, u_I^H, \dots, u_{N/2}^H)$  describes a *displacement* of the fine-grid function  $u^h = (u_0, \dots, u_i, \dots, u_N)$ ; i.e., it modifies the latter through *interpolation and addition*:

$$u^h = \tilde{u}^h + I_H^h u^H \tag{11a}$$

where  $\tilde{u}^h$  is the fine-grid configuration at the stage of switching to the coarse grid and  $I_H^h$  denotes the following *weighted* interpolation from grid  $H$  to grid  $h$ :

$$(I_H^h u^H)_i = \begin{cases} u_I^H & \text{if } i = 2I \\ \frac{a_i u_I^H + a_{i+1} u_{I+1}^H}{a_i + a_{i+1}} & \text{if } i = 2I + 1 \end{cases} \tag{11b}$$

(see Sec. 2.4 for the reasons behind this prescription).

The fine-grid Hamiltonian  $\mathcal{H}_h(u^h)$  resulting from that interpolation can be written as follows:

$$\mathcal{H}_h(\tilde{u}^h + I_H^h u^H) = \mathcal{H}_h(\tilde{u}^h) + \mathcal{H}_H(u^H) \quad (12)$$

where  $\mathcal{H}_h(\tilde{u}^h)$  is given by (10) and  $\mathcal{H}_H(u^H)$  is:

$$\mathcal{H}_H(u^H) = \frac{1}{H} \sum_{I=1}^{N/2} a_I^H (u_I^H - u_{I-1}^H)^2 + H \sum_{I=1}^{N/2-1} \phi_I^H u_I^H \quad (13)$$

with

$$a_I^H = \frac{2a_{i-1}a_i}{a_{i-1} + a_i} \quad (I = i/2 = 1, 2, \dots, N/2) \quad (14a)$$

and

$$\begin{aligned} \phi_I^H = & -\frac{a_I^H}{2h^2} \tilde{u}_{i-2} + \frac{a_I^H + a_{I+1}^H}{2h^2} \tilde{u}_i - \frac{a_{I+1}^H}{2h^2} \tilde{u}_{i+2} \\ & + \frac{a_i}{2(a_{i-1} + a_i)} \phi_{i-1} + \frac{\phi_i}{2} + \frac{a_{i+1}}{2(a_{i+1} + a_{i+2})} \phi_{i+1} \\ & (I = i/2 = 1, 2, \dots, N/2 - 1) \end{aligned} \quad (14b)$$

The coarse-grid couplings  $a_I^H$  depend only on the fine-grid couplings  $a_i^h$ . The coarse field terms  $\phi_I^H$  are calculated from the details of the fine-grid configuration at coarsening and are fixed throughout the processing on the coarser level. The variables of the coarse grid  $u_I^H$  are initially set to zero, corresponding to zero initial displacements.

Having calculated the field  $\phi^H$  once for all,  $\mathcal{H}_H$  can now be directly calculated in terms of the coarse grid configuration  $u^H$ ; there is no need to explicitly perform (12) in order to perform a Monte Carlo step on the coarser level. One can therefore run a long Monte Carlo process with  $\mathcal{H}_H$  before explicitly updating  $u^h$  by (11a).

The Monte Carlo process for  $\mathcal{H}_H$  can itself include a transition to a still coarser grid,  $2H$ , and so on. Thus, more precisely, the entire algorithm is defined as a sequence of multigrid cycles for the *finest* level, where a cycle for any given (“*current*”) level is *recursively* defined by the following five stages.

1.  $v_i^1$  Monte Carlo *steps* are first made for each variable  $u_i$  on the current level.  $v_i^1 \geq 0$  and may change from one cycle to another. In the case of constant coupling,  $v_i^1$  will generally also be a constant



( $v_i^1 = v^1$ ), and then the steps are actually performed as a sequence of  $v^1$  Monte Carlo *sweeps*, each including one step per gridpoint. Similar organization in *partial sweeps* is suitable for *variable*  $v_i^1$ .

2. If the current level is the coarsest, go to 5. Otherwise, the next coarser level is created from the current one by determining its couplings (14a) and field terms (14b).
3.  $\gamma$  multigrid cycles for the coarse level are performed. The “cycle index”  $\gamma$  may change from one current level to another.
4. Update the current level by performing (11a).
5. Additional  $v_i^2$  Monte Carlo *steps* are finally made at each variable  $u_i$  on the current level.  $v_i^2 \geq 0$  and may change from one cycle to another, and the steps may well again be ordered in (partial) sweeps.

The Monte Carlo steps are performed by changing each variable in its turn randomly according to its associated distribution, regarding its neighbors as fixed.

The values of  $v_i^1$ ,  $v_i^2$  and  $\gamma$  are discussed below.

Exactly the same five stages define also the multigrid cycle in a general domain  $\Omega$  in any dimension. For the recursion, the  $d$ -dimensional Hamiltonian in (5) is generalized to

$$\mathcal{H}_h(u^h) = h^{d-2} \sum_{\langle i, j \rangle} a_{ij}^h (u_i^h - u_j^h)^2 + h^d \sum_i \phi_i^h u_i^h \quad (15)$$

where  $\sum_i$  runs over all interior sites. Since the interpolation  $I_{2h}^h$  is a linear operator, the coarse grid Hamiltonians will again have the same form as (15), except that the range of neighbors  $i$  for each site  $j$  (i.e., the points  $i$  for which  $a_{ij}^{2h} \neq 0$ ) will depend on the order of the weighted interpolation (see Sec. 2.4). It will be explained below that the couplings at the fine grid  $h$  determine the coefficients of the interpolation operator  $I_{2h}^h$ . As a result,  $a_{ij}^{2h}$  will depend on the couplings at grid  $h$ , and the field  $\phi_i^{2h}$  is directly derived from the current fine grid configuration  $\tilde{u}^h$ ,  $a_{ij}^h$  and from  $\phi_i^h$ , similar to ((14a)–(14b)) in the one-dimensional case. In our standard  $d$ -dimensional example we take  $\Omega = [0, L]^d$ ,  $h = L/N$ ,  $N = 2^k$ , so that  $h^d/|\Omega| = N^{-d}$ .

## 2.4. Weighted Interpolation

Why the particular form (11b) is chosen for the interpolation operator  $I_{2h}^h$ ? For efficient coarse-grid sampling, all physically probable large-scale configuration changes should have approximations of the form  $I_{2h}^h u^{2h}$ , with

approximately the same energy changes. For any probable set  $u^{2h}$  of (probable) displacements at the subset of points belonging to the coarse grid,  $I_{2h}^h u^{2h}$  should give a similarly probable set of displacements at all points. Otherwise, large-scale movements  $u^{2h}$  would be associated with energy differences much larger than physically probable, prohibiting their amplitudes from approaching physically probable sizes, yielding inefficient coarse grid sampling: all likely samples would remain in some neighborhood of the original configuration  $\tilde{u}$ .

For example (see Fig. 1), if  $a_i \gg a_{i+1}$  at some fine grid site  $i = 2I + 1$  (not belonging to the coarse grid), then  $u_i$  is likely to be much closer to  $u_{i-1}$  than to  $u_{i+1}$ , hence the usual linear interpolation, if used instead of (11b), would enforce unlikely moves, unless  $u_i^H - u_{i+1}^H$  is small. Thus, under linear interpolation, likely coarse-grid moves would have small differences, hence a small overall amplitude (see demonstrations in Sect. 2.7 below).

Given any neighboring values  $u_{i-1}^h$  and  $u_{i+1}^h$ , the most probable value for  $u_i^h$  is that which minimizes (10), satisfying

$$\frac{\partial \mathcal{H}_i^h(u)}{\partial u_i^h} = \frac{2a_i}{h} (u_i^h - u_{i-1}^h) - \frac{2a_{i+1}}{h} (u_{i+1}^h - u_i^h) + h\phi_i = 0 \quad (16)$$

Hence, when the displacements in  $u_{i-1}^h$  and  $u_{i+1}^h$  are  $u_{i-1}^H$  and  $u_{i+1}^H$ , respectively, the most likely displacement in  $u_i^h$  is given by (11b), since this keeps (16) unchanged.

Similar considerations can be used to derive weighted interpolation for general *higher-dimensional* problems (except that, for obtaining relations analogous to (16), the Hamiltonian should first be restricted to a local set of points around the point  $i$ , and its minimization for any given coarse-grid values should be done over all non-coarse points in the set, yielding a certain value at  $i$  which is a function of the coarse values). More simply, except in certain pathological cases, one can generally use compounded forms of (11b). For example, the weighted interpolation  $I_{2h}^h$  that we have

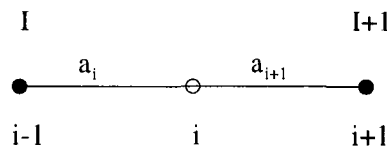


Fig. 1. One-dimensional grid: the fine grid couplings are denoted by  $a_i$  and  $a_{i+1}$ , and the bolded points are the coarse grid points.

used for the two-dimensional problem is given in terms of the notations in Fig. 2, by

$$\begin{aligned}
 (I_{2h}^h u^{2h})_r &= \begin{cases} u_I^{2h} & \text{if } r = i \\ u_O^{2h} & \text{if } r = o \\ u_Q^{2h} & \text{if } r = q \\ u_K^{2h} & \text{if } r = k \\ \frac{a_{ij}u_I^{2h} + a_{jk}u_K^{2h}}{a_{ij} + a_{jk}} & \text{if } r = j \\ \frac{a_{il}u_I^{2h} + a_{lo}u_O^{2h}}{a_{il} + a_{lo}} & \text{if } r = l \\ \frac{a_{op}u_O^{2h} + a_{pq}u_Q^{2h}}{a_{op} + a_{pq}} & \text{if } r = p \\ \frac{a_{kn}u_K^{2h} + a_{nq}u_Q^{2h}}{a_{kn} + a_{nq}} & \text{if } r = n \\ \frac{a_{mj}(I_{2h}^h u^{2h})_j + a_{ml}(I_{2h}^h u^{2h})_l + a_{mp}(I_{2h}^h u^{2h})_p + a_{mn}(I_{2h}^h u^{2h})_n}{a_{mj} + a_{ml} + a_{mp} + a_{mn}} & \text{if } r = m \end{cases} \\
 & \hspace{15em} (17)
 \end{aligned}$$

where  $a_{ij}$  is the fine grid coupling between sites  $i$  and  $j$ . Other slightly different definitions for  $I_{2h}^h$  could also be used.<sup>(1, 4, 11, 13, 15)</sup> (Prescription (17) results from choosing the smallest possible local set in deriving each of the relations analogous to (16). In pathological cases larger sets should be used.)

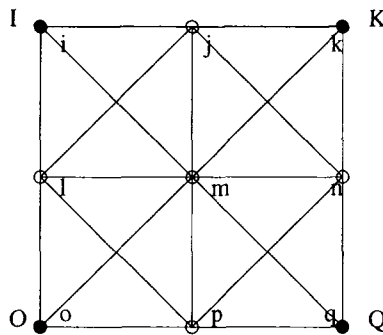


Fig. 2. Two-dimensional grid: the fine grid connections are denoted by solid lines, the bolded points are the coarse grid points.

## 2.5. Analysis for Fast Sampling of Susceptibility

In case of constant-coupling Gaussian models it has been shown<sup>(7,9,10,16)</sup> that the susceptibility is dominated by contributions from large-scale fluctuations (low-frequency Fourier Components). Therefore, the purpose of the simulation is to sample quickly as many such fluctuations as possible. An optimal multigrid algorithm is achieved by applying a cycle index  $\gamma$  larger than  $2^d$  and calculating the susceptibility by averaging over many measurements on the coarsest level.

This is *not* always the exact situation in the variable-coupling Gaussian models. Here, the scales that dominate the contribution to the susceptibility are determined by the strength of the coupling  $a(x)$ . An optimal Monte Carlo process should sample more frequently regions with smaller values of the coupling  $a(x)$ .

In the *constant*-coupling models, the Fourier components are mutually independent. Moreover, since the Monte Carlo process is local, a relaxation sweep on a certain level changes effectively only those Fourier components with wavelength comparable to the meshsize of that level. These two observations enabled us<sup>(7,9,10,16)</sup> to estimate the number of Monte Carlo sweeps needed at different levels of the multigrid cycle. In the *variable*-coupling Gaussian model the Fourier analysis no longer holds. Instead, we develop a new type of analysis that approximately decouples the various scales. In this analysis each configuration is written as a combination of local movements from all levels. These movements will not be entirely independent (see e.g., condition (21) below), but they will suffice for the purpose of our considerations.

The analysis and our numerical experiments (see Sec. 2.6) treat zero boundary conditions, but the derived rules will be suitable for general boundary conditions.

The magnetization  $M_h$  can be evaluated on any level, without going back to finer levels (plug (11a)–(11b) into (6) to obtain an expression of  $M_H$  as an explicit linear function of  $u_H$ ). Thus, with negligible extra work, many measurements of  $M_h^2$  can be made within a cycle, and their average can be used as an estimate for the discrete susceptibility  $\langle M_h^2 \rangle$ .

To study the number  $n_{ki}$  of relaxation *steps* that the algorithm needs to perform at site  $i$  on a grid with meshsize  $h_k = 2^k h$ , ( $k=0, 1, \dots, l = \log_2(N/2)$ ), in order to achieve accuracy  $\varepsilon$  in the estimation of the susceptibility, let the number of internal sites at level  $k$  be denoted by  $v_k$ , and let  $u_i^k = u^k(x_i^k)$  be the spin at  $x_i^k$  (site  $i$  on level  $k$ ), where  $i \in V_k = \{j = (j_1, \dots, j_d) : 1 \leq j_1, \dots, j_d \leq N/2^k - 1\}$ . The coarsest level ( $k=l$ ) includes only one internal variable.

The interpolation from any level  $k$  to the next-finer level  $k-1$  is denoted by  $I_{h_k}^{h_{k-1}}$  or simply  $I_k^{k-1}$ . It is defined by (11b), where  $H$  is

substituted with  $h_k, h$  with  $h_{k-1}$  and the couplings  $a_i$  are those of level  $k-1$ .

We define a multiscale set of basis functions as follows. Each level  $k$  is associated with  $v_k$   $\delta$ -functions  $\{\delta^{i,k}\}_{i \in V_k}$  defined on that level by

$$\delta^{i,k}(x_j^k) = \delta_{ij} \quad (j \in V_k) \tag{18}$$

Our set of basis functions are then defined as the *fine-grid* functions

$$\beta^{i,k} = I_1^0 I_2^1 \dots I_k^{k-1} \delta^{i,k} \tag{19}$$

Each fine-grid configuration  $u^0$  can then be represented as

$$u^0(x_j^0) = \sum_{k=0}^l \sum_{i \in V_k} c_i^k \beta^{i,k}(x_j^0) \quad (j \in V_0) \tag{20}$$

where the coefficients  $\{c_i^k\}_{i \in V_k}^{k=0, \dots, l}$  are uniquely determined by requiring the total contribution to (20) of each finer level ( $k-1$ ) to be orthogonal to each basis function of the next-coarser level ( $k$ ); i.e., requiring

$$\sum_{j \in V_{k-1}} (I_k^{k-1} \delta^{i,k})_j c_j^{k-1} = 0 \quad (k = 1, \dots, l \quad i \in V_k) \tag{21}$$

Indeed, to see that the coefficients  $c_j^k$  are uniquely determined, note first that their number ( $\sum_{k=0}^l v_k$ ) equals the number of equations in (20)–(21). Then note that (21) implies the orthogonality relations

$$\left( \sum_{k=0}^{m-1} \sum_{j \in V_k} c_j^k \beta^{j,k}, \beta^{i,m} \right) = 0 \quad (m = 1, \dots, l \quad i \in V_m) \tag{22}$$

Hence, for a given configuration  $u^0$ , the single coarsest-level coefficient  $c_1^l$  is determined by

$$c_1^l(\beta^{1,l}, \beta^{1,l}) = (u^0, \beta^{1,l}) \tag{23}$$

Furthermore, having determined all the coefficients  $c_i^k$  for all  $k > m$ , a set of  $v_m$  equations for the level- $m$  coefficients  $\{c_j^m\}_{j \in V_m}$  is given by

$$\sum_{j \in V_m} (\beta^{j,m}, \beta^{i,m}) c_j^m = (u^{0,m}, \beta^{i,m}) \quad (i \in V_m) \tag{24}$$

where

$$u^{0,m} = u^0 - \sum_{k=m+1}^l \sum_{i \in V_k} c_i^k \beta^{i,k}$$

from which the coefficients  $\{c_j^m\}_{j \in V_m}$  are uniquely determined since the  $v_m \times v_m$  matrix  $(\beta^{j,m}, \beta^{i,m})$  is strongly diagonally dominant. Therefore, the representation (20) is indeed unique.

This representation enables us to understand exactly the role of each step in the multigrid cycle. A relaxation step at  $x_i^k$  mainly changes  $c_i^k$ , while  $c_i^k$  for any  $j \neq i$  is much less affected. A relaxation sweep on level  $k$  effectively samples  $\{c_j^k\}_{j \in V_k}$ , the stochastic coefficients of that level, while its effect on any other  $c_i^m (m \neq k)$  is drastically reduced as  $m - k$  increases, and completely vanishes for  $m < k$  (due to the orthogonality relations). See details of particular cases demonstrated in Appendix A.

According to (20) the magnetization (6) can be expressed as a linear combination of the stochastic variables  $\{c_i^k\}_{i \in V_k}^{k=0, \dots, l}$ , namely,

$$M_h(u) = \sum_{k=0}^l \sum_{i \in V_k} s_i^k c_i^k \quad (25)$$

Hence

$$M_h^2(u) = \sum_{k=0}^l \sum_{m=0}^l \sum_{i \in V_k} \sum_{j \in V_m} s_{ij}^{km} c_i^k c_j^m \quad (26)$$

where  $s_{ij}^{km} = s_i^k s_j^m$ . Since we are interested in the estimation of the susceptibility

$$\chi = \langle M^2 \rangle - \langle M \rangle^2 = \langle M^2 \rangle = \lim_{h \rightarrow 0} \langle M_h^2 \rangle \quad (27)$$

to a certain accuracy  $\varepsilon$ , the number  $n_{ki}$  of relaxation steps at each  $u_i^k$  depends on the standard deviation of each term in (26), namely on

$$s_{ij}^{km} \sigma(c_i^k c_j^m) = s_{ij}^{km} (\langle (c_i^k c_j^m)^2 \rangle - \langle c_i^k c_j^m \rangle^2)^{1/2} \quad (28)$$

Since  $\langle c_i^k \rangle = 0$  for any  $k=0, \dots, l$  and  $i \in V_k$ , and  $c_i^k, c_j^m$  are almost uncorrelated for  $(i, k) \neq (j, m)$  (see Appendix A), it follows that

$$s_{ij}^{km} \sigma(c_i^k c_j^m) \approx s_{ij}^{km} \sigma(c_i^k) \sigma(c_j^m) \quad (29)$$

Now, we make an arbitrary partition of  $\{c_i^k\}$  into  $R$  subsets:  $(i, k)$  and  $(j, m)$  will be in the same subset  $I_r$  if and only if  $s_i^k \sigma(c_i^k) \approx s_j^m \sigma(c_j^m)$ . In other words, for all  $(i, k) \in I_r$  we assume  $s_i^k \sigma(c_i^k) \approx \sigma_r$ . Then, for each  $(i, k) \in I_r$  we require our algorithm to satisfy  $n_{ki} \geq n_r$ , where  $n_r$  increases with the typical standard deviation  $\sigma_r$  of the subgroup.

If  $(i, k)$  and  $(j, m)$  are in the same subgroup  $I_r$ , then the term  $s_{ij}^{km} c_i^k c_j^m$  in (26) is sampled at least  $n_r$  times. For a given  $r$ , there are  $|I_r|^2$  such terms

which are only weakly correlated (see Appendix A), therefore, using (29), the standard deviation of

$$\sum_{(i,k),(j,m) \in I_r} s_{ij}^{km} c_i^k c_j^m$$

is at most

$$\frac{|I_r| \sigma_r^2}{\sqrt{n_r}} \tag{30}$$

If  $(i, k) \in I_r$  and  $(j, m) \in I_t$  and  $\sigma_r > \sigma_t$ , then  $s_{ij}^{km} c_i^k c_j^m$  is sampled at least  $n_r$  times. Therefore, for a given  $r$  and  $t$  with  $\sigma_r > \sigma_t$ , there are  $|I_r| |I_t|$  such terms which are almost uncorrelated (see Appendix A), hence the standard deviation of

$$\sum_{(i,k) \in I_r} \sum_{(j,m) \in I_t} s_{ij}^{km} c_i^k c_j^m$$

is at most

$$\frac{\sqrt{|I_r| |I_t|} \sigma_r \sigma_t}{\sqrt{n_r}} \tag{31}$$

One can combine the error estimations in (30) and (31) and get the total error estimate in measuring the susceptibility  $\langle M^2 \rangle$

$$\varepsilon \approx \sum_r \frac{\sqrt{|I_r|} \sigma_r}{\sqrt{n_r}} \sum_{t: \sigma_t \leq \sigma_r} \sqrt{|I_t|} \sigma_t \tag{32}$$

We assume that the deviation in each configuration is dominated by the largest typical standard deviation (largest  $\sigma_r$ ), i.e., the quantity  $\sqrt{|I_r|} \sigma_r$  decreases geometrically with  $r$ . Therefore, the total error estimate in measuring the susceptibility and the total number of computer operations are respectively

$$\varepsilon \approx \sum_r \frac{|I_r| \sigma_r^2}{\sqrt{n_r}} \tag{33}$$

and

$$W \approx \sum_r |I_r| n_r \tag{34}$$

The optimal choice for  $n_r$  (yielding either minimal  $\varepsilon$  for a given  $W$  or minimal  $W$  for a given  $\varepsilon$ ) is obtained when  $\partial\varepsilon/\partial n_r + \lambda_1 \partial W/\partial n_r = 0$ , which by (33) and (34), yields

$$n_r = \lambda_2 \sigma_r^{4/3} \quad (35)$$

where  $\lambda_1$  and  $\lambda_2$  are independent of  $r$ . Indeed, insertion of (35) into (33) and (34) yields the optimal relation  $W \approx \varepsilon^{-2}$ , as long as  $\sum_r |I_r| \sigma_r^{4/3}$  is bounded independently of  $N^d$ . Since clearly  $\varepsilon = O(\sigma)$  for  $W = O(1)$ , where  $\sigma$  is the standard deviation of the susceptibility, another way to write the obtained relations is  $W = O(\sigma^2 \varepsilon^{-2})$ .

It is important to emphasize that the result in (35) is independent of  $|I_r|$ . Henceforth, changing the partitioning into subsets will not change the result. This *optimal variable sampling rule* can also be written as

$$n_{ki} = O((s_i^k \sigma(c_i^k))^{4/3}) \quad (36)$$

In order to construct an optimal multigrid algorithm with a convenient sampling rule, it is necessary to approximate  $s_i^k \sigma(c_i^k)$  for  $k = 0, \dots, l$  and  $i \in V_k$ . The contribution extent of  $c_i^k$  to the magnetization,  $s_i^k$ , can be computed directly from (24), giving

$$s_i^k \approx \frac{h^d}{L^d} 2^{kd} = \frac{h^d}{L^d} \quad (k = 0, \dots, l \quad i \in V_k) \quad (37)$$

According to the observations in Appendix A

$$\sigma(c_i^k) = O\left(\left(\frac{h_k^{2-d}}{A_i^k}\right)^{1/2}\right) \quad (k = 0, \dots, l \quad i \in V_k) \quad (38)$$

where  $A_i^k = \sum_{j: \langle i, j \rangle} a_{ij}^k$  denotes the sum of the couplings extending from site  $i$  on level  $k$ . Substitution of (37) and (38) into (36) yields the sampling rule that we have used in our numerical experiments

$$n_{ki} = C(A_i^k)^{-2/3} h_k^{(4+2d)/3} \quad (39)$$

where  $C$  is a constant independent of  $k$  and  $i$ .

This rule implies the cycle index

$$\gamma = 2^{(4+2d)/3} \quad (40)$$

since, by (14a), values of  $A_i^k$  on different levels (different  $k$ ) are comparable. However, the number of Monte Carlo steps actually taken at each site of



each visited level is governed by the more precise rule (39) (see examples in Sec. 2.6).

We conclude with a remark concerning the discretization error. The size  $N^d$  of the finest grid that should be employed increases of course with the decrease of  $\varepsilon$ , because one needs to have a grid for which the computed average is only distance  $\varepsilon$  from its infinite-grid value. Therefore, it is necessary to construct a discretization scheme, for the observable in question, with the dependence  $N = N(\varepsilon)$  such that  $N(\varepsilon)^d \leq O(\varepsilon^{-2})$ . In other words, one should apply a discretization scheme with an error smaller than  $O(1/\sqrt{N^d})$ , i.e., a discretization of order at least  $d/2$ . (We have used a second-order discretization, for  $d=1$  and  $d=2$ .)

## 2.6. Numerical Results

We have tested the multigrid algorithm, by applying the variable sampling rule (39), for different coupling functions in one and two dimensional lattices with grids of sizes up to 1024 and 1024<sup>2</sup> respectively. Our main aim was to show that using the variable sampling rule properly the susceptibility can be calculated in an optimal-time. The susceptibility has been measured over just *one* cycle. Within the cycle, many measurements of the magnetization  $M_h$  are taken, in fact after each relaxation sweep on each level. The average of the measurements,  $\overline{M_h^2}$ , is an approximation for  $\langle M_h^2 \rangle$  (7), which is also an approximation for the thermodynamic limit  $\langle M^2 \rangle$  (4). The *relative accuracy* is defined as  $\varepsilon_r = |\overline{M_h^2} - \langle M^2 \rangle|/\sigma$ , where  $\sigma$  denotes the standard deviation of the susceptibility. We define the *performance index*  $\alpha$  to be the expected value of  $\# \text{RAN} \cdot \varepsilon_r^2$ , where  $\# \text{RAN}$  is the amount of work spent in the cycle, measured by the number of times a random number is generated. Thus,  $\alpha$  should turn out to be bounded if and only if the algorithm solves to relative accuracy  $\varepsilon_r$  in  $O(\varepsilon_r^{-2})$  operations, or in other words, the algorithm eliminates completely both the critical slowing down and the volume factor. We measured  $\alpha$  for different kinds of step functions and coupling strengths. Results are presented below. In most cases  $\varepsilon_r$  is averaged over an ensemble of 10<sup>4</sup> runs; for the cases that smaller ensembles were used, the deviation in measuring  $\alpha$  is given in parantheses.

**Example 1.** In the 1D problem on  $\Omega = [0, 1]$ , the couplings are defined by the step function

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.5 \\ a_* & \text{if } 0.5 \leq x \leq 1 \end{cases}$$

Table II. Performance Indices  $\alpha$  for Example 1

$a_*$	$N$								
	4	8	16	32	64	128	256	512	1024
1	2.10	3.22	3.70	3.97	4.09	4.25	3.99(0.18)	4.03(0.18)	4.50(0.20)
10	2.41	4.36	5.79	6.65	6.85	7.57	7.63(0.31)	7.31(0.33)	8.07(0.35)
1000	2.71	4.08	5.45	6.21	6.66	6.93	7.35(0.31)	6.64(0.30)	7.10(0.31)
10000	0.64	1.71	2.78	3.56	3.98	4.45	4.35(0.19)	4.29(0.19)	4.39(0.20)

For large  $a_*$ , the subinterval  $[0.5, 1]$  is strongly coupled, hence the configuration there is almost a constant. This subinterval should therefore be relaxed much less than the subinterval  $[0, 0.5]$ . Practically, according to the variable sampling rule, the number of relaxation sweeps over  $[0.5, 1]$  should be  $1/a_*^{2/3}$  times their number over  $[0, 0.5]$ , and each level must be relaxed 4 times as many sweeps as the next finer level, except that the coarsest level ( $H = 1/2$ ), where the weak couplings happen to completely disappear is visited very rarely: only once per  $a_*^{2/3}$  cycles. The resulting values of  $\alpha$  are presented in Table II,<sup>2</sup> for different  $a_*$ . It is clear that, independently of the coupling strength  $a_*$ , an optimal efficiency comparable to the constant-case ( $a_* = 1$ ) efficiency is obtained, i.e.,  $\alpha$  remains uniformly bounded as  $N$  grows.

**Example 2.** Again on  $\Omega = [0, 1]$ , the couplings are now defined by the step function

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.25 \\ a_* & \text{if } 0.25 \leq x \leq 0.75 \\ 1 & \text{if } 0.75 < x \leq 1 \end{cases}$$

The number of relaxation steps over the subinterval  $[0.25, 0.75]$  should by (39) be  $1/a_*^{2/3}$  times the number elsewhere. On the coarsest level, the strong couplings are no longer present, turning this level to be by far the most sampled one. As before, cycle index 4 has been used. The measurements of  $\alpha$  are presented in Table III, demonstrating again optimal efficiency. Note that for this example, pointwise Monte Carlo schemes (see Sec. 2.2) would yield  $\alpha = O(a_* N^3)$ !

<sup>2</sup> For the constant-coupling case, uniform sampling with cycle index  $\gamma = 4$  is used.

Table III. Performance Indices  $\alpha$  for Example 2

$a_*$	$N$								
	4	8	16	32	64	128	256	512	1024
1	2.10	3.22	3.70	3.97	4.09	4.25	3.99(0.18)	4.03(0.18)	4.50(0.20)
10	1.35	2.22	2.70	3.08	3.13	3.26	3.96(0.18)	3.96(0.17)	4.25(0.19)
1000	0.95	1.17	1.56	1.79	1.98	2.04	2.14(0.09)	2.06(0.09)	2.08(0.09)
10000	0.94	1.19	1.56	1.84	1.96	2.02	1.99(0.09)	2.12(0.09)	2.11(0.09)

Example 3. The coupling function is defined by

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.1875 \\ a_* & \text{if } 0.1875 \leq x \leq 0.4375 \\ 1 & \text{if } 0.4375 < x < 0.5625 \\ a_* & \text{if } 0.5625 \leq x \leq 0.8125 \\ 1 & \text{if } 0.8125 < x \leq 1 \end{cases}$$

In this example, the steps (jumps) in the coupling function coincides with the geometry of the subintervals only for grids with meshsize  $H \leq 1/16$ . As in the previous examples, cycle index 4 has been employed with suitable variable sampling according to (39). Note that on a grid with meshsize  $H = 1/8$  the strong couplings still hold, but on the two coarsest grids the strong couplings do not exist anymore. Hence, on these two grids uniform sampling (full sweep) should be done. The measurements of  $\alpha$  are presented in Table IV, demonstrating optimal efficiency, independently of  $a_*$ .

Table IV. Performance Indices  $\alpha$  for Example 3

$a_*$	$N$						
	16	32	64	128	256	512	1024
1	3.70	3.97	4.09	4.25	3.99(0.18)	4.03(0.18)	4.50(0.20)
100	3.01	3.29	3.53	3.50	3.26(0.14)	3.32(0.16)	3.29(0.14)
1000	2.91	3.24	3.34	3.44	3.56(0.15)	3.56(0.16)	3.24(0.15)

Example 4. In the 2D problem on  $\Omega = \{0 \leq x, y \leq 1\}$ , the couplings are defined by the step function

$$a(x, y) = \begin{cases} a_* & \text{if } 0.25 \leq x, y \leq 0.75 \\ 1 & \text{otherwise} \end{cases}$$

Table V. Performance Indices  $\alpha$  for Example 4

$a_*$	$N^2$								
	$4^2$	$8^2$	$16^2$	$32^2$	$64^2$	$128^2$	$256^2$	$512^2$	$1024^2$
1	3.63	8.71	13.46	16.58	18.84	20.38(0.41)	22.91(1.07)	20.42(0.93)	21.18(1.00)
1000	1.05	2.49	4.56	6.24	7.55	8.84(0.42)	12.83(2.01)	13.88(2.14)	11.53(1.45)

Here, cycle index 6 has been used, accompanied with rule (39). The results are presented in Table V,<sup>3</sup> demonstrating again optimal behavior ( $\alpha$  practically bounded independently of  $N$  and  $a_*$ ). By contrast, the pointwise Monte Carlo would yield  $\alpha = O(a_* N^4)$ .

## 2.7. Two-Level Diagnostic Tests

The two-level tests enable us to check and better understand the performance of the multigrid algorithm. In this kind of test we estimate the deviation from the desired observable average (e.g., the discrete susceptibility  $\langle M_h^2 \rangle$ ) introduced by each coarsening from some fine level (meshsize  $h$ ) to the next-coarser level (meshsize  $2h$ ). Here, we will use such an analysis to check the performance of different interpolation schemes. The examples are such that weighted interpolation is required already on the fine grid. For each transition from a fine-grid equilibrium configuration to the coarse grid we have measured the discrete susceptibility after each relaxation sweep on the coarse grid, and calculated the difference between the average of very many (practically infinite) such measurements and the true fine-grid discrete susceptibility  $\langle M_h^2 \rangle$ ; this difference represents the deviation caused by the fine-to-coarse transition. Averaging this deviation over an ensemble of many fine-grid equilibrium configurations yields the “average coarsening deviation” (ACD). Note that if the number of passes on the coarse grid is not sufficiently large (as in some of our experiments below) the measured ACD may be much larger than the true one.

**Example 1.** The couplings are defined by the step function

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.25 \\ 1000 & \text{if } 0.25 \leq x \leq 0.75 \\ 1 & \text{if } 0.75 < x \leq 1 \end{cases}$$

for  $N=4$  and  $h=1/4$ . The standard deviation is  $\sigma_h=0.0497$ .

<sup>3</sup> For the constant-coupling case, uniform sampling with cycle index  $\gamma=6$  is used.

**Table VI. Average Coarsening Deviation in the Susceptibility for Example 1**

Number of passes on the coarse grid	ACD using linear interpolation	ACD using weighted interpolation
1000	0.049	0.00158
2000	0.048	0.00111
4000		0.000777
8000		0.000560
$\infty$	0.052	0.0000300

Table VI presents the coarsening deviation in susceptibility, averaged over an ensemble of  $4 \cdot 10^5$  configurations in the case of linear interpolation and over  $4 \cdot 10^3$  configurations in the case of weighted interpolation.

In this example, the coarse grid includes only one internal point, thus the statistics on that level can be computed analytically ( $\infty$  passes). Generally, in problems that require much sampling at one site in a certain level, the statistics accumulation can be accelerated by analytical averaging at that site.

The coarsening deviation in the case of linear interpolation has the same order as the standard deviation  $\sigma_h$ . In the case of weighted interpolation the true ACD is smaller by a factor of about  $1.6 \cdot 10^3$  (essentially proportional to  $a_*$ ). When the number  $n$  of passes (and effective measurements) on the coarse grid is not sufficiently large, an additional deviation of about  $\sigma_h/\sqrt{n}$  enters the measured ACD.

We conclude that an optimal algorithm can be constructed with weighted interpolation and *not* with linear interpolation.

**Example 2.** The couplings are defined by the step function

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.375 \\ 1000 & \text{if } 0.375 \leq x \leq 0.625 \\ 1 & \text{if } 0.625 < x \leq 1 \end{cases}$$

for  $N = 8$  and  $h = 1/8$ . The standard deviation is  $\sigma_h = 0.0574$ .

Table VII presents the coarsening deviation in susceptibility, averaged over ensemble of  $4 \cdot 10^3$  configurations in the case of weighted interpolation and over  $10^5$  configurations in the case of linear interpolation (except for 32000 and 64000 passes where the ensemble sizes are 30000 and 2000, respectively).

Table VII. Average Coarsening Deviation in the Susceptibility for Example 2

Number of passes on the coarse grid	ACD using linear interpolation	ACD using weighted interpolation
1000	0.0378	0.00339
2000	0.0298	0.00266
4000	0.0226	0.00209
8000	0.0172	0.00183
16000	0.0133	0.00159
32000	0.0109	0.00151
64000	0.0096	

In the previous example, the linear interpolation weights and the weighted interpolation weights were totally different over the whole interval, resulting in large ratio between the linear interpolation ACD and the weighted interpolation ACD. Here, the linear interpolation coincides with the weighted interpolation except for the weights at the fine grid points 0.375 and 0.625. Indeed, the average coarsening deviation using weighted interpolation is smaller only by a factor of 6 compared with that of the linear interpolation. More “jumps” in the variable-coupling function would demonstrate much worse performance of the linear interpolation relative to that of the weighted interpolation.

### 3. SUMMARY

The calculation of a thermodynamic limit of any observable to a relative accuracy  $\varepsilon_r$  by a usual Monte Carlo process requires  $O(N^{d+z}\varepsilon_r^{-2})$  computer operations, where  $\varepsilon_r$  is the error relative to the standard deviation of the observable,  $N$  is the linear dimension of the lattice needed to approximate the thermodynamic limit to accuracy  $\varepsilon_r$ ,  $d$  is the dimension and  $z$  is the critical exponent.

In the variable-coupling isotropic Gaussian models the overall work might increase as

$$O(a_* N^{d+z}\varepsilon_r^{-2})$$

where  $a_*$  denotes the maximal ratio between values of the coupling.

Multigrid algorithms can reduce and even eliminate not only the *critical slowing down* factor  $a_* N^z$  but also the *volume factor*  $N^d$ .

By a novel method, the parameters of the multigrid algorithm such as the cycle index  $\gamma$  and the sampling rule can be determined as functions of the coupling coefficients. For the optimal calculation of the susceptibility in

the variable-coupling Gaussian model it is essential to use weighted interpolation and the variable sampling rule.

The optimal efficiency is obtained *independently* of the coupling function discontinuities, with performance as good as in the constant-coupling case. The critical slowing down *and* the volume factor are completely eliminated, and the total required computational work is just  $O(\varepsilon_r^{-2})$ .

### APPENDIX A: NUMERICAL EXAMPLES FOR ILLUSTRATING THE ANALYSIS

To illustrate the details of the theoretical analysis (Sec. 2.5), some numerical examples are provided below. We consider the one-dimensional variable-coupling Gaussian model with three different coupling functions. For each example we first solve the related system ((20) and (21)) and then measure the following three properties:

- (a) The typical local amplitude  $\langle (c_i^k)^2 \rangle^{1/2}$  for  $k = 0, \dots, l$  and  $i \in V_k$ .
- (b) The change introduced to any  $c_i^0$  as a result of one relaxation step on the finest grid, relative to  $\langle (c_i^0)^2 \rangle^{1/2}$ .
- (c) The average change introduced to  $c_i^k$  as a result of one relaxation sweep on the finest grid, relative to  $\langle (c_i^k)^2 \rangle^{1/2}$ . This average is defined as

$$\frac{1}{v_k} \sum_{i \in V_k} \frac{(\sum_m ((c_i^k)_{m-1} - (c_i^k)_m)^2)^{1/2}}{M^{1/2} \langle (c_i^k)^2 \rangle^{1/2}}$$

where  $m$  is the relaxation sweep number and  $M = 10^5$  is the ensemble size (number of sweeps).

**Example 1.** The constant-coupling case,  $a(x) \equiv 1$  for  $x \in [0, 1]$ .

The typical local amplitude:  $\langle (c_i^k)^2 \rangle^{1/2} = O(h_k^{1/2})$ .

A relaxation step at  $x_i^0$  changes mostly  $c_i^0$ , while  $c_j^0$  for  $j \neq i$  is much less affected: Fig. 3 presents an example of the influence of one relaxation step at  $x_i^0$  on  $c_j^0$ , showing a drastic decrease as  $j$  moves away from  $i$ .

Table VIII shows the average relative change introduced to  $c_i^k$  as a result of one relaxation sweep on the finest level (ensemble size is  $M = 10^5$ ). The fast weakening of the effect with increasing  $k$  is apparent.

Similarly; a relaxation sweep on any level  $m$  would introduce large relative changes on that level, with the changes in  $c_j^k$  decreasing exponentially as function of  $k - m$  if  $k > m$ , and with no change at all in  $c_j^k$  if  $k < m$ .

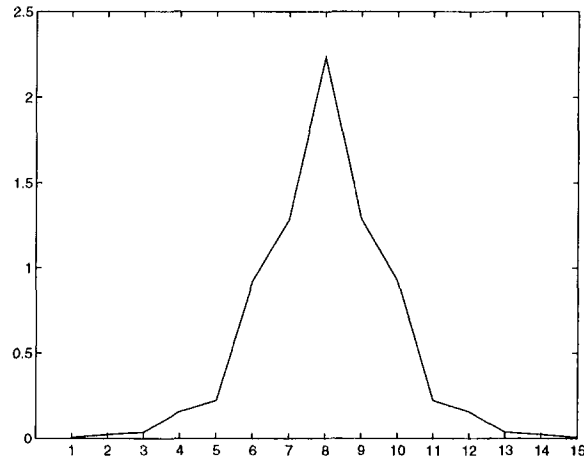


Fig. 3. The change of  $c_j^0$  relative to  $\langle (c_j^0)^2 \rangle^{1/2}$ , as a result of one relaxation step at  $x_g^0$ , with  $N=16$ .

A corollary of all these observations is the weak correlation between different levels, i.e.,  $\{c_i^k\}_{i \in V_k}$  are weakly correlated to  $\{c_i^m\}_{i \in V_m}$  for  $k \neq m$ . Moreover in order to sample effectively all scales, a multilevel algorithm should be introduced (see Sec. 2.5).

**Table VIII. Measuring the Change Introduced to  $c_i^k$ , Relative to  $\langle (c_i^k)^2 \rangle^{1/2}$ , As a Result of One Relaxation Sweep on the Finest ( $k=0$ ) Level**

$N$	$k=0$	$k=1$	$k=2$	$k=3$
4	1.42	0.91		
8	1.44	0.98	0.52	
16	1.45	1.01	0.60	0.27

**Example 2.** The 1D variable coupling

$$a(x) = \begin{cases} a_* = 100 & \text{if } 0 \leq x \leq 0.5 \\ 1 & \text{if } 0.5 < x \leq 1 \end{cases}$$

As in the constant-coupling case,  $c_i^k$  represents the local amplitude of the configuration on level  $k$  at site  $i$ , with weak correlation between local



**Table IX. Measuring the Change Introduced to  $c_i^k$ , Relative to  $\langle (c_i^k)^2 \rangle^{1/2}$ , As a Result of One Relaxation Sweep on the Finest Level**

$N$	$k=0$	$k=1$	$k=2$	$k=3$
4	1.41	1.29		
8	1.41	0.96	0.87	
16	1.43	1.00	0.60	0.53

amplitudes on the same level. While the typical local amplitude in the weakly coupled subdomain is  $\langle (c_i^k)^2 \rangle^{1/2} = O(h_k^{1/2})$ , the typical local amplitude in the strongly coupled subdomain is  $\langle (c_i^k)^2 \rangle^{1/2} = O(h_k^{1/2}/\sqrt{a_*})$ . Therefore, using (36), the weakly coupled subdomain should be sampled about  $a_*^{2/3}$  times as often as the strongly coupled subdomain.

Table IX shows the relative change in  $c_i^k$  as a result of one relaxation sweep on the finest level (ensemble size  $M=10^5$  for  $N=4$  and 8, and  $M=2 \cdot 10^5$  for  $N=16$ ).

The general behavior exhibited in Table IX is very much the same as in the constant-coefficient case (Table VIII). It might also be noted that the variable  $s_3^{l-1} c_3^{l-1}$  turns out in this example to have the largest standard deviation, which practically means that the site  $x_3^{l-1}$  should be the most sampled site.

**Example 3.** The 1D variable coupling

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq x < 0.25 \\ a_* = 100 & \text{if } 0.25 \leq x \leq 0.75 \\ 1 & \text{if } 0.75 < x \leq 1 \end{cases}$$

**Table X. Measuring the Change Introduced to  $c_i^k$ , Relative to  $\langle (c_i^k)^2 \rangle^{1/2}$ , As a Result of One Relaxation Sweep on the Finest Level**

$N$	$k=0$	$k=1$	$k=2$	$k=3$
4	1.37	0.13		
8	1.42	1.32	0.16	
16	1.42	0.97	0.90	0.11

As in the cases mentioned above, the deviations at all scales have been found to have weak correlations between them. The typical local amplitudes are  $\langle (c_i^k)^2 \rangle^{1/2} = O(h_k^{1/2})$  and  $\langle (c_i^k)^2 \rangle^{1/2} = O(h_k^{1/2}/\sqrt{a_*})$  in the weak-coupled domain and in the strong-coupled domain, respectively.

In this case the Monte Carlo process suffers severe slowness. The measurements, presented in Table X (ensemble size is  $M = 2 \cdot 10^5$ ), confirm this observation. The variable  $s_1^l c_1^l$  has the largest standard deviation, but the change introduced to the movement on that coarsest grid is only a small fraction (roughly  $a_*^{-1/2}$ ) of the typical movement.

## APPENDIX B: MODIFICATIONS TO THE ANISOTROPIC CASE

The general  $d$ -dimensional anisotropic Hamiltonian is defined in the continuum by

$$\mathcal{H}(u) = \int_{\Omega} \left( \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 u}{\partial x^i \partial x^j} \right) dx^1 \dots dx^d$$

where the matrix  $\{a_{ij}(x)\}$  is positive definite at every point  $x$ . Even when the continuum problem is isotropic, the problem emerging at some coarser level of the multigrid algorithm may well have severe anisotropies.

A measure of the anisotropy at a point  $x$  is given by the ratio

$$\alpha = \frac{\max \sum_{ij} a_{ij}(x) \xi_i \xi_j}{\min \sum_{ij} a_{ij}(x) \xi_i \xi_j}$$

where both the max and the min are taken over the sphere  $\sum_{i=1}^d \xi_i^2 = 1$ . Experience and theory developed in multigrid solvers for anisotropic PDE problems<sup>(2,3)</sup> indicate that the following modifications to the basic algorithm are necessary under severe anisotropy (large  $\alpha$ ), and are useful already at milder anisotropy ( $\alpha \geq 3$ , say).

In case the anisotropy is consistently aligned with the discretization grid, exhibiting uniformly strong couplings in some direction and uniformly weak in others, for instance

$$a_{11} \approx a_{22} \approx \dots \approx a_{kk} \gg a_{k+1,k+1} \approx a_{k+2,k+2} \approx \dots \approx a_{dd}$$

and

$$a_{ij}^2 \ll a_{ii} a_{jj}$$

then *semi-coarsening* in the strong-coupling directions ( $x_1, x_2, \dots, x_k$ ) should be employed. That is, the next coarser grid in the multigrid algorithm

should have twice the fine-grid meshsize only in those (strong-coupling) directions, while in all other directions the meshsize should remain as in the fine grid.

In more general situations, an approach similar to “algebraic multigrid” (AMG) can be adopted. In AMG solvers for PDE,<sup>(4, 11, 25)</sup> the next-coarse-level variables are typically selected by the requirement that each current-fine-level variable is “strongly connected” to at least some coarse-level variables in its neighborhood.

## ACKNOWLEDGMENTS

The research has been supported in parts by grants No. G0289-065.07/93 from the German-Israeli Foundation for Research and Development (GIF), No. 379/93 from the Israeli Academy of Science and Humanities and by the Carl F. Gauss Minerva Center for Scientific Computation.

## REFERENCES

1. R. E. Alcouffe, A. Brandt, J. E. Dendy, Jr., and J. W. Painter: The diffusion equation for strongly discontinuous coefficients, *SIAM J. Sci. Stat. Comp.* **2**:430–454 (1981).
2. A. Brandt, *Math. Comp.* **31**:333 (1977).
3. A. Brandt, Multigrid techniques: 1984 Guide, with applications to Fluid Dynamics (available as GMD Studien Nr. 85, GMD-AIW, Postfach 1240, D-5205, St. Augustin 1, Germany).
4. A. Brandt, Algebraic multigrid theory: the symmetric case, *Appl. Math. Comp.* **19**:23–56 (1986).
5. A. Brandt, Multilevel Computations: Reviews and Recent Developments, in *Preliminary Proc. 3rd Copper Mountain conference on Multigrid Methods* (April 1987). See also in *Multigrid Methods: theory applications and super-computing*, S. F. McCormick, ed. (Marcel Dekker, 1988), pp. 35–62.
6. A. Brandt, The Weizmann Institute Research in Multilevel Computation: 1988 Report, in: *Proceedings 4th Copper Mountain Conf. on Multigrid Methods*, J. Mandel et al., eds. SIAM (1989) pp. 13–53.
7. A. Brandt, Multigrid methods in lattice field computations, *Nucl. Phys. B (Proc. Suppl.)* **26**:137–180 (1992).
8. A. Brandt and M. Galun, The anharmonic crystal model, in preparation.
9. A. Brandt and M. Galun, Optimal multigrid algorithms for the massive Gaussian model and path integrals, *J. Stat. Phys.* **82**:1503–1518 (1996).
10. A. Brandt, M. Galun, and D. Ron, Optimal multigrid algorithms for calculating thermodynamic limits, *J. Stat. Phys.* **74**:313–348 (1994).
11. A. Brandt, S. McCormick, and J. Ruge, algebraic multigrid theory (AMG) for automatic multigrid solutions with application to geodetic computations, 1982 Report, Institute for computational studies, Fort Collins, CO.
12. A. Brandt, D. Ron, and D.J. Amit, Multi-level approaches to discrete-state and stochastic problems, in *Multigrid Methods*, W. Hackbusch and U. Trottenberg, eds. (Springer Verlag, 1986), pp. 66–99.

13. J. E. Dendy, Jr., *J. Comp. Phys.* **48**:366 (1982); *Appl. Math. Comp.* **13**:261 (1983); *Appl. Math. Comp.* **19**:57 (1986); *Appl. Math. Comp.* **25**:1 (1988).
14. R. G. Edwards, J. Goodman, and A. D. Sokal, Multi-grid Monte Carlo II, Two-dimensional XY model, *Nucl. Phys. B* **354**:289 (1991).
15. R. G. Edwards, S. J. Ferreira, J. Goodman, and A. D. Sokal, Multi-grid Monte Carlo III, Two-dimensional  $O(4)$ -symmetric non-linear  $\sigma$ -model, *Nucl. Phys. B* **380**:621–664 (1992).
16. M. Galun, Optimal multigrid algorithms for model problems in statistical mechanics, M.Sc. Thesis, Weizmann Institute of Science (1992).
17. J. Goodman and A. D. Sokal, Multigrid Monte Carlo methods for lattice field theories, *Phys. Rev. Lett.* **56**:1015–1018 (1986).
18. M. Hasenbusch, S. Meyer, and G. Mack, Noncritical multigrid Monte Carlo:  $O(3)$  non-linear  $\sigma$  model, *Nucl. Phys. B (Proc. Suppl.)* **20**:110–113 (1991).
19. D. Kandel, E. Domany, and A. Brandt, Simulations without critical slowing down—Ising and 3-State Potts models, *Phys. Rev.* **B40**:330 (1989).
20. D. Kandel, E. Domany, D. Ron, A. Brandt, and E.-Loh, Jr., Simulations without critical slowing down, *Phys. Rev. Lett.* **60**:1591 (1988).
21. G. Mack and A. Pordt, Convergent perturbation expansions for euclidean quantum field theory, *Comm. Math. Phys.* **97**:267 (1985); G. Mack, in *Nonperturbative Quantum Field Theory*, G. t'Hooft *et al.* (Eds.), Plenum Press, NY, 309 (1988).
22. J. Machta, Y.S. Choi, A. Lucke, T. Schweizer, and L. V. Chayes, Invaded cluster algorithm for equilibrium critical points, *Phys. Rev. Lett.* **75**:2792 (1995).
23. G. Mana, T. Mendez, A. Pelissetto, and A. D. Sokal, Dynamic critical behavior of multigrid Monte Carlo for two-dimensional nonlinear  $\sigma$ -models, to appear in *Nucl. Phys. B (Proc. Suppl.)*.
24. T. Mendez and A. D. Sokal, One-dimensional  $O(4)$ -symmetric nonlinear  $\sigma$ -model, to appear in *Phys. Rev. D*.
25. J. Ruge and K. Stüben, Algebraic multigrid, in *Multigrid methods*, S. McCormick (ed.), Frontiers in applied mathematics, SIAM 5 Philadelphia (1987).
26. R. H. Swendsen and J. S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
27. U. Wolff, Collective Monte Carlo updating for spin systems, *Phys. Rev. Lett.* **62**:361–364 (1989).