Statistically Optimal Multigrid Algorithms for the Anharmonic Crystal Model

A. Brandt & M. Galun

Dept. of Applied Mathematics & Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel

Abstract

Two types of multigrid algorithms for the one dimensional anharmonic crystal model are presented. The first type applies linear interpolation operators and the second type applies nonlinear interpolation operators with approximate Hamiltonians on coarse grids. For both algorithms, the question of eliminating the "volume" complexity factor is examined, i.e., the feasibility of the algorithm to remove the need to produce many independent fine-grid configurations for averaging out their statistical deviations, so that thermodynamic limits can be calculated to relative accuracy ε_r in just $O(\varepsilon_r^{-2})$ computer operations, where ε_r is the error relative to the standard deviation of the observable. The main difficulty arising in the nonlinear anharmonic crystal model is the coupling between different scales. In this paper, it is shown by analysis and numerical tests that the multigrid algorithm with the linear interpolation operators can eliminate the volume factor only partially, i.e., upto a certain accuracy. A multigrid algorithm with nonlinear interpolation operators that can eliminate the volume factor completely is described in details and tested experimentally.

KEY WORDS: multigrid; Monte Carlo; critical slowing down; volume factor; statistically optimal algorithm; thermodynamic limit; anharmonic crystal model; approximate minimization interpolation operator.

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 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 [6,13,19] can eliminate the CSD (i.e., obtain $z = 0$). For more complicated models, (e.g. ϕ^4, nonlinear σ-models or discrete models) more recent publications report on simulation techniques that partially [7,8,9,13,20,21,22,24] or completely [1,14,15,16,18,25] 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.

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 ε_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 ε_r is the error relative to σ, the standard deviation of the observable in question. Also, the size N^d of the grid must increase as some positive power of ε_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, although there exists a coupling between different scales, we will show below that by suitable cycling and interpolation procedures one can completely remove both the volume factor and the CSD.

The simultaneous elimination of the volume factor and the CSD factor means that a thermodynamic limit can be calculated to an accuracy of $\pm\varepsilon$ in *"statistically optimal time"*, i.e. in only $O(\varepsilon_r^{-2}) = O(\sigma^2 \varepsilon^{-2})$ *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. An algorithm that can calculate a thermodynamic limit to an accuracy of $\pm\varepsilon$ in statistically optimal time is called *"statistically optimal algorithm"*. In other words, a statistically optimal algorithm effectively produces an independent sample in just $O(1)$ computer operations. 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 [2,5,11] 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 ε_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 ε_r in about $20\varepsilon_r^{-2}$ random number generations. Then, we have shown for the one dimensional *massive* Gaussian model [3] that the susceptibility is calculated to relative accuracy ε_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. Moreover, results as good as those previously obtained [2,3,5,11] for constant coefficients, were attained [4] for the non-constant coupling Gaussian models with strongly discontinuous cases in one and two dimensions. For the one dimensional variable-coupling Gaussian model, the susceptibility is calculated to accuracy ε_r in less than $8\varepsilon_r^{-2}$ random number generations. In the two-dimensional variable-coupling Gaussian model, the susceptibility can be measured in less than $20\varepsilon^{-2}$ random generations. These results are independent of the maximal ratio between the values of the coupling, unlike the severe extra slowness that large such ratios can inflict on conventional Monte Carlo [4].

The present paper treats the one dimensional *anharmonic* crystal model. The Hamiltonian of the anharmonic crystal model is a fourth order polynomial, unlike the quadratic polynomial Hamiltonians we have treated so far [4,3,5,11]. As a result of the nonlinearity of the anharmonic crystal model, a new major difficulty is presented: coupling between different scales. This difficulty is common to all advanced models, e.g. $O(N)$ and $SU(N)$ models. However, the anharmonic crystal model is simpler than the later, since it has no topologies. Therefore, the new problem of coupling between different scales can be studied without struggling with other new difficulties, such as the appearence of topologies on various scales.

While Fourier expansions were used for analyzing the constant-coefficient Gaussian model and the massive Gaussian model, they could no longer serve the variable-coupling Gaussian model, neither for exact calculations of continuum and discrete averages, nor for analyzing the multigrid Monte-Carlo simulations. Therefore, we developed a new approximate analysis that helped us to construct a statistically optimal multigrid algorithm for the variable-coupling Gaussian models. This novel analysis (with a certain improvement introduced in the present paper) provides us with an important tool for understanding the nonlinear anharmonic crystal model, where non-constant couplings stochastically emerge at coarser levels of the multigrid Monte Carlo processing. It is shown by this analysis, as well as experimentally that multigrid algorithms with *linear interpolation operators* can eliminate the volume factor partially (upto a certain accuracy) in measuring the susceptibility. However, in order to obtain truly statistically optimal multigrid

algorithm in the one dimensional anharmonic crystal model, one should apply appropriate *nonlinear interpolation operators*, which clearly demand *approximations* to the Hamiltonians on coarse grids. The way to construct a multigrid algorithm with the nonlinear interpolation operators, so that the volume factor is completely eliminated, is described in details, in Secs. 6 and 7, below, and numerical tests are reported in Sec. 8.

2. Continuum and discrete model

The *continuous* Hamiltonian of the *one-dimensional* anharmonic crystal model is

$$\mathcal{H}(u) = \int_0^L (\lambda u_x^4 + u_x^2) dx, \tag{1}$$

where $u = u(x)$ is defined for $x \in [0, L]$, subscripts stand for derivatives, and $\lambda \geq 0$. Boundary conditions $u(0) = u(L) = 0$ are assumed. The simple Gaussian model is obtained when $\lambda = 0$. The first term in the Hamiltonian is the *anharmonic* oscillator and the second term is the *harmonic* oscillator.

Discrete approximations will be calculated by placing a grid of points $x_i = ih$ over the domain $[0, L]$, where $i = 0, 1, \ldots, N$ and $h = L/N$ being the meshsize. The value of the discrete configuration u^h at the point x_i will be denoted u_i^h or u_i and $u_0 = u_N = 0$ will express the boundary conditions. The *discrete* Hamiltonian of the one dimensional anharmonic crystal model will be given by

$$\mathcal{H}_h(u^h) = \frac{\lambda}{h^3} \sum_{i=1}^{N} (u_i - u_{i-1})^4 + \frac{1}{h} \sum_{i=1}^{N} (u_i - u_{i-1})^2. \tag{2}$$

The probability density of the configurations in the continuous case is the Boltzmann distribution

$$P(u) = \frac{1}{Z} \exp^{-\mathcal{H}(u)/T} \tag{3}$$

where T is the temperature and Z is a normalization factor (the partition function) derived from the condition $\int_u P(u)du = 1$. In the discrete case, the probability distribution is given again by (3), with $\mathcal{H}_h(u^h)$ replacing $\mathcal{H}(u)$.

Usually, the discrete magnetization is defined as

$$M_h(u) = \frac{h}{L} \sum_i u_i$$

where the summation is over all internal sites. As the density function (3) is given, average properties of interest are the average discrete magnetization $\langle M_h \rangle = \int_u M_h(u) P(u) du$ and the discrete susceptibility

$$\chi_h = \langle M_h^2 \rangle - \langle M_h \rangle^2.$$

Clearly, in the case of the sign-symmetric Hamiltonian (2) and the homogeneous boundary conditions, $\langle M_h \rangle = 0$.

We will see later that, with this conventional definition of the discrete susceptibility, its thermodynamic limit vanishes, i.e., $\chi_h \to 0$ as $h \to 0$. Therefore, a proper scaling should be introduced. Henceforth, one of the problems is to find an appropriate scaling factor $\frac{h^a}{L^a}$ such that, upon re-defining

$$M_h(u) = \frac{h^a}{L^a} \sum_i u_i \tag{4}$$

and

$$\chi_h = \langle M_h^2(u) \rangle = \frac{h^{2a}}{L^{2a}} \langle (\sum_i u_i)^2 \rangle \tag{5}$$

the thermodynamic limit for the susceptibility (5) would neither vanish nor diverge as $h \to 0$.

3. The multigrid algorithm with linear interpolation operators

In this section, we describe a multigrid algorithm whose interpolation operators are *linear* (see definition below). Later on, by applying the multilevel approximate analysis method that we have introduced firstly for the variable-coupling Gaussian model [4], we explain (Sec. 4.1), and confirm by actual simulations, that upto a certain accuracy, performance similar to that of the Gaussian model can still be obtained for the anharmonic model (see Ex. 3 in Sec. 8), although this requires a careful choice of the multigrid cycling parameters (Sec. 5). Beyond that accuracy we will need a more advanced interpolation (Sec. 6).

For recursive purposes we will consider a somewhat generalized anharmonic crystal model in the interval $[0, L]$, whose Hamiltonian $\mathcal{H}_h(u^h)$ on a grid with meshsize $h = L/N = L/2^k$ includes non-constant couplings:

$$\mathcal{H}_h(u) = \sum_{i=1}^{N} a_i (u_i - u_{i-1} + b_i)^4 + \sum_{i=1}^{N} c_i (u_i - u_{i-1} + d_i)^2. \tag{6}$$

On the *finest* grid we will assume that the special form (2) is given, i.e., $a_i = \frac{\lambda}{h^3}$, $b_i = 0$, $c_i = \frac{1}{h}$ and $d_i = 0$ $(1 \le i \le N)$.

3.1. Coarse grid displacement

The coarse grid with meshsize $H = 2h$ is constructed by taking every other grid-point. The coarse-grid function $u^H = (u_0^H, \ldots, u_I^H, \ldots, u_{N/2}^H)$ is interpreted as

a displacement of the fine grid function $u^h = (u_0, \ldots, u_i, \ldots, u_N)$, updating the latter through interpolation and addition:

$$u^h = \tilde{u}^h + I_H^h u^H, \tag{7}$$

where $\tilde{u}^h$ is the previous fine grid configuration, i.e., the one existing just before switching to the coarse grid, and I_H^h denotes a *linear* interpolation operator from grid H to grid h, having the general form

$$(I_H^h u^H)_i = \sum_{I=1}^{N/2-1} w_{iI} u_I^H \qquad (i = 1, \ldots, N-1).$$

By calling this interpolation *linear* we mean that each weight w_{iI} can be a function only of $\tilde{u}^h$ (which is fixed throughout the simulation on level H), but not a function of the coarse grid configuration u^H. The choice of w_{iI} will usually be such that $\sum_{I=1}^{N/2-1} w_{iI} = 1$ for any $i = 1, \ldots, N-1$, and part (usually most) of them will be equal to zero. When the non-zero weights are *not* equal to each other, we say that the linear interpolation operator is a *weighted* linear interpolation operator.

Here, we applied the following weighted linear interpolation operator from grid H to grid h:

$$\left(I_H^h u^H\right)_i = \begin{cases} u_I^H & \text{if } i = 2I \\ w_i u_I^H + w_{i+1} u_{I+1}^H & \text{if } i = 2I+1, \end{cases} \tag{8}$$

where

$$w_i = \frac{\sqrt{a_i} + c_i}{\sqrt{a_i} + c_i + \sqrt{a_{i+1}} + c_{i+1}}; \quad w_{i+1} = \frac{\sqrt{a_{i+1}} + c_{i+1}}{\sqrt{a_i} + c_i + \sqrt{a_{i+1}} + c_{i+1}} \tag{9}$$

(see Sec 3.3 for the reason behind this choice of weights).

The updated fine-grid Hamiltonian can be separated into two parts as follows:

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

where $\mathcal{H}_h(\tilde{u}^h)$ is given by (6) and the coarse grid Hamiltonian $\mathcal{H}_H(u^H)$ is given by

$$\mathcal{H}_H(u^H) = \sum_{I=1}^{N/2} a_I^H (u_I^H - u_{I-1}^H + b_I^H)^4 + \sum_{I=1}^{N/2} c_I^H (u_I^H - u_{I-1}^H + d_I^H)^2 \tag{10}$$

with

$$a_I^H = a_i w_{i-1}^4 + a_{i-1} w_i^4, \tag{11}$$

$$b_I^H = \frac{a_i w_{i-1}^3(\tilde{u}_i - \tilde{u}_{i-1} + b_i) + a_{i-1} w_i^3(\tilde{u}_{i-1} - \tilde{u}_{i-2} + b_{i-1})}{a_i w_{i-1}^4 + a_{i-1} w_i^4}, \tag{12}$$

$$\begin{aligned} c_I^H = & \frac{6 a_{i-1} a_i w_{i-1}^2 w_i^2 (w_i(\tilde{u}_i - \tilde{u}_{i-1} + b_i) - w_{i-1}(\tilde{u}_{i-1} - \tilde{u}_{i-2} + b_{i-1}))^2}{a_i w_{i-1}^4 + a_{i-1} w_i^4} \\ & + c_i w_{i-1}^2 + c_{i-1} w_i^2 \end{aligned} \tag{13}$$

and

$$\begin{aligned} d_I^H = & \frac{1}{2 c_I^H} (4 a_i w_{i-1} (\tilde{u}_i - \tilde{u}_{i-1} + b_i)^3 + 4 a_{i-1} w_i (\tilde{u}_{i-1} - \tilde{u}_{i-2} + b_{i-1})^3 \\ & + 2 c_i w_{i-1} (\tilde{u}_i - \tilde{u}_{i-1} + d_i) + 2 c_{i-1} w_i (\tilde{u}_{i-1} - \tilde{u}_{i-2} + d_{i-1}) \\ & - \frac{4 (a_i w_{i-1}^3 (\tilde{u}_i - \tilde{u}_{i-1} + b_i) + a_{i-1} w_i^3 (\tilde{u}_{i-1} - \tilde{u}_{i-2} + b_{i-1}))^3}{(a_i w_{i-1}^4 + a_{i-1} w_i^4)^2}), \end{aligned} \tag{14}$$

for $(I = i/2 = 1, \ldots, N/2 - 1)$.

The coarse-grid couplings a_I^H and c_I^H and the coarse "field" terms b_I^H and d_I^H are calculated from the fine-grid couplings and from the fine-grid configuration at coarsening and are fixed throughout the processing on the coarser level. Thus, a long Monte Carlo process can be done on the coarser level, with the Hamiltonian $\mathcal{H}_H$, without explicitly updating u^h by (7). The variables of the coarse grid u_I^H are initially set to zero, i.e., no displacement.

The process of calculating a^H, b^H, c^H and d^H, then simulating $\mathcal{H}_H(u^H)$ by a Monte Carlo process on the coarse level, then interpolating the resulting u^H to the fine grid and changing u^h by (7), is called a *coarse grid displacement.*

Since the coarse grid Hamiltonian, $\mathcal{H}_H(u^H)$, has the same general structure (10) as the fine grid Hamiltonian $\mathcal{H}_h(u^h)$, its Monte Carlo process can itself include a *coarser* grid (grid $2H$) displacements. Thus, the coarsening can be employed recursively, using increasingly coarser grids (see Sec. 3.4)

The interpolation from any level k to the the next-finer level $k-1$, denoted I_k^{k-1}, is defined by (8) and (9), where H is substituted with h_k, h with h_{k-1} and the couplings a_i and c_i are those of level $k-1$. Note that due to the specific form of the finest grid Hamiltonian (2), it turns out that I_1^0 is a pure (non-weighted) linear interpolation, i.e., all the weights are equal to $1/2$.

3.2. Heat-bath relaxation

For simulating the generalized anharmonic crystal Hamiltonian on any level, a heat-bath relaxation is used. To implement the heat-bath simulation at any point, we need [12] a subroutine that generates a random variable X from the probability distribution with density proportional to

$$f(x) = e^{-a_4 x^4 - a_3 x^3 - a_2 x^2 - a_1 x}$$

where $a_4 > 0$ and a_1, a_2, a_3 are arbitrary. By a shift in x it can be assumed that $a_3 = 0$.

A von Neumann rejection algorithm is used [17]: given a Gaussian density function $g(x) \geq f(x)$, a random variable X with density proportional to $g(x)$ is generated and then accepted with probability $f(X)/g(X)$, keeping trying until success. The acceptance fraction is

$$A = \frac{\int_{-\infty}^{\infty} f(x)dx}{\int_{-\infty}^{\infty} g(x)dx}$$

and the expected number of trials is A^{-1}.

In our case the Gaussian distribution

$$g(x) = e^{-\alpha_2 x^2 - \alpha_1 x - \alpha_0} \quad \alpha_2 > 0$$

is used, its parameters chosen subject to the constraint $g(x) \geq f(x)$ so as to maximize the acceptance fraction A. It turns out that $\alpha_1 = a_1$ and $\alpha_0 = -\frac{(\alpha_2 - a_2)^2}{4a_4}$, where α_2 is the unique solution of a cubic equation, which can be solved easily by a few iterations of Newton's method.

Another possible approach of course is to use the Metropolis relaxation [23] (see Sec. 6.2).

3.3. Weighted interpolation

If one would like to apply a *linear* interpolation operator, then in order to produce probable configurations effectively, suitably weighted interpolation must be used, similarly to the weighted interpolation in the variable-coupling Gaussian models [4]. Here, the weight on bond i should be proportional to $\sqrt{a_i^h} + c_i^h$, for the following reason.

Given any neighboring values u_{i-1}^h and u_{i+1}^h, the most probable value for u_i^h at some fine grid site $i = 2I + 1$ (not belonging to the coarse grid) is that which minimizes (6), satisfying

$$\begin{aligned} \frac{\partial \mathcal{H}_h(u)}{\partial u_i^h} &= 4a_i(u_i^h - u_{i-1}^h + b_i)^3 + 2c_i(u_i^h - u_{i-1}^h + d_i) \\ &\quad - 4a_{i+1}(u_{i+1}^h - u_i^h + b_{i+1})^3 - 2c_{i+1}(u_{i+1}^h - u_i^h + d_{i+1}) = 0. \end{aligned} \tag{15}$$

Therefore, when the displacements in u^h_{i-1} and u^h_{i+1} are $\delta_{i-1} = u^H_I$ and $\delta_{i+1} = u^H_{I+1}$, respectively, then in order to keep equation (15) unchanged the most likely displacement δ_i in u^h_i should satisfy

$$\begin{aligned} &6a_i(u^h_i - u^h_{i-1} + b_i)^2(\delta_i - \delta_{i-1}) + c_i(\delta_i - \delta_{i-1}) \\ -&6a_{i+1}(u^h_{i+1} - u^h_i + b_{i+1})^2(\delta_{i+1} - \delta_i) - c_{i+1}(\delta_{i+1} - \delta_i) = 0 \end{aligned} \tag{16}$$

where higher-order displacement *differences* are neglected since we should mainly be interested in *smooth* coarse-grid displacements.

If on bonds i and $i+1$ the anharmonic part in (6) dominates then statistically $a_i(u^h_i - u^h_{i-1} + b_i)^2 \propto \sqrt{a_i}$ and $a_{i+1}(u^h_{i+1} - u^h_i + b_{i+1})^2 \propto \sqrt{a_{i+1}}$, hence according to (16) the most likely displacement δ_i in u^h_i is given by (8) and (9). If on bonds i and $i+1$ the Gaussian part in (6) dominates then as in the variable-coupling Gaussian models w_i should be proportional to c_i. Practically, the weights which are given in (9) would cover both cases.

Since on the finest grid the couplings a^h_i and c^h_i are positive, according to (11) and (13) the coarse grid couplings a^H_I and c^H_I are also positive. Thus, the interpolation weights (9) are well defined and positive.

3.4. The compound multigrid cycle

The entire algorithm can be described 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. ν_1 Monte Carlo sweeps are first made on the current level.
2. If the current level is the coarsest, goto 5. Otherwise, the next coarser level is created from the current one by determining its couplings (11) and (13) and field terms (12) and (14).
3. γ multigrid cycles for the coarse level are performed. *The "cycle index" γ may change from one current level to another and need not be an integer* (see below).
4. Update the current level by performing the weighted interpolation (7) from the coarse level.
5. Additional ν_2 Monte Carlo sweeps are finally made on the current level.

Let γ_k be the cycle index for level k ($k = 0, \ldots, l = \log_2(N/2)$). To implement cycle indices which are not integers, the approximate maintenance of the following condition is used to determine whether to make an additional cycle from any intermediate level k:

$$nc(k \to k+1) \approx \gamma_k \cdot nc(k-1 \to k)$$

where $nc(j \to j+1)$ counts the total number of transitions that have been made so far from a fine grid j to the next coarser grid $j+1$.

Since the weighted interpolation is a linear operator, the magnetization M_h (and hence the susceptibility) can be evaluated on any level, without going back to finer levels (plug (7) into M_h 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 $\langle M_h^2 \rangle$. Practically, measurements are taken only on the coarsest level, after each relaxation sweep there, because only there substantial changes in M_h are introduced.

Now, in order to choose the optimal multigrid cycling parameters (ν_1, ν_2 and γ), it is first necessary to analyze the coupling between different scales in the susceptibility. This is done in the following section (Sec. 4). Then, we explain why the multigrid algorithm with linear interpolation operators cannot achieve statistical optimality at too-high levels of accuracy, as a certain amount of coarsening bias is introduced into the susceptibility estimation (Sec. 4.1). Thereafter, we describe the optimal choice of the multigrid cycling parameters (Sec. 5).

4. Analysis of the coupling between different scales in the susceptibility

In the anharmonic crystal model, as in the simple Gaussian model [5] and in the massive Gaussian model [3,11], the susceptibility is still dominated by the smoothest component. Unlike the Gaussian models, however, the size of this component is strongly affected here by other components, especially by the high-frequency ones. In this section, we estimate the amplitudes of typical fluctuations at various scales and the effect of the small-scale fluctuations on the susceptibility.

In order to understand the model multiscale characteristics, we use the multilevel approximate analysis method introduced earlier for the variable-coupling Gaussian model [4], with a certain improvement introduced in the present work. Level k ($k = 0, 1, \ldots, l = \log_2(N/2)$) is associated with a grid with meshsize $h_k = 2^k h$. Let the number of internal sites at level k be denoted by $v_k = N/2^k - 1$, and let V_k denote the set $\{1, 2, \ldots, v_k\}$ and $\bar{V}_k = \{1, 2, \ldots, v_k, v_k + 1\}$. The variable at $x_i^k = ih_k$ (site i on level k) is denoted by $u_i^k = u^k(x_i^k)$. The coarsest level ($k = l$) includes only one internal variable u_1^l ($v_l = 1$).

The interpolation from any level k to the next finer level $k-1$, denoted by I_k^{k-1}, is the above *weighted linear interpolation operator* (see Sec. 3.1).

We use for our analysis (following Ref. 4) a *multiscale* set of basis functions. Each level k is associated with v_k δ–functions $\{\delta^{i,k}\}_{i \in V_k}$ defined on that level by

$$\delta^{i,k}(x_j^k) = \delta_{ij} \qquad (j \in V_k).$$

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

$$\beta^{i,k} = I_1^0 I_2^1 \cdots I_k^{k-1} \delta^{i,k}.$$

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

$$u^0(x_j^0) = \sum_{k=0}^{l} \sum_{i \in V_k} c_i^k \beta^{i,k}(x_j^0) \qquad (j \in V_0) \tag{17}$$

where the coefficients $\{c_i^k\}_{i \in V_k}^{k=0,\ldots,l}$ are uniquely determined by requiring

$$c_{2i}^k = 0 \qquad (k = 0, \ldots, l-1) \tag{18}$$

i.e., on any level k each even-index coefficient is zero. Note that in Ref. 4we have used a different requirement, based on orthogonality relations between any finer level $(k-1)$ and the next coarser level (k). The requirement (18) can also be expressed as an orthogonality condition, but with respect to the dot product

$$\langle u, v \rangle = \sum_i w_i (u_i - u_{i-1})(v_i - v_{i-1})$$

which yields more convenient relations.

This representation enables us to understand the role of each step in the multigrid cycle. A level-k relaxation step at x_{2i+1}^k changes only c_{2i+1}^k. A level-k relaxation step at x_{2i}^k, of course would not change $c_{2i}^k \equiv 0$, but the coarser level coefficient c_i^{k+1} would be affected, if i is odd, and a still-coarser level value will be affected if i is even. However, those coarser level values are mainly sampled by relaxation at their level. Actually, a relaxation *sweep* on level k effectively samples all c_{2j+1}^k, the stochastic coefficients of that level, its effect on any other c_{2i+1}^m completely vanishes for $m < k$, while for $m \geq k$ it is only $O(4^{-(m-k)})$ relative to the typical variation of that coefficient. Therefore, each stochastic variable c_{2i+1}^k can be regarded as a local amplitude at site $2i+1$ on level k. Moreover, the representation of u^0 by (17) allows us to express the magnetization as a linear combination of stochastic variables $\{c_i^k\}_{i \in V_k}^{k=0,\ldots,l}$, such that the susceptibility is strongly dominated by $(c_1^l)^2$, the next main contributions being from $\{c_i^{l-1}\}_{i \in V_{l-1}}$, etc.

Having defined the local amplitudes $\{c_i^k\}_{i \in V_k}^{k=0,\ldots,l}$ as above, it will be natural to define a *local slope* s_i^k on level k $(k = 0, \ldots, l)$ by

$$s_i^k = \frac{c_i^k - c_{i-1}^k}{h_k} \qquad (i \in \bar{V}_k).$$

Note that due to (18)

$$s^k_{2i-1} = -s^k_{2i} \qquad k = 0, \ldots, l.$$

By introducing u^0 (as represented in (17)) into the Hamiltonian (2) we will now obtain estimations for the local amplitudes and for the local slopes.

First we show that large-scale levels, being smoother, contribute less to $\mathcal{H}_h(u)$. For that purpose, we express the Hamiltonian (2) in local slope terms. One can show that for the above weighted linear interpolation operator, it can be assumed that

$$\beta^{i,k}(x^0_j) - \beta^{i,k}(x^0_{j-1}) = O(2^{-k}) \qquad (j \in \bar{V}_0) \tag{19}$$

for any $k = 0, \ldots, l$ and $i \in \bar{V}_k$. (Note that for $k = 0, 1$ the right-hand-side of (19) is $\pm 2^{-k}$ or 0). By plugging the fine-grid configuration representation (17) into (2), and by using the local slope definition and the estimation in (19), one would obtain the Hamiltonian in slope terms. It can be easily shown that the Hamiltonian in slope terms should contain terms of the form $\lambda h_k (s^k_i)^4$ and $h_k (s^k_i)^2$. Therefore, statistically,

$$\max[h_k (s^k_i)^2, \lambda h_k (s^k_i)^4] \leq O(T),$$

yielding

$$s^k_i \leq O(\min[T^{1/2} h_k^{-1/2}, \lambda^{-1/4} T^{1/4} h_k^{-1/4}]) \qquad (k = 0, \ldots, l). \tag{20}$$

For $k = 0$ we must actually have equality, i.e.,

$$s^0_i = O(\min[T^{1/2} h^{-1/2}, \lambda^{-1/4} T^{1/4} h^{-1/4}]) \tag{21}$$

otherwise the Hamiltonian terms involving s^0_i would not add up to an $O(T)$ total value.

Since the order of magnitude of s^0_i does not depend on the slope position, we use the following notation without the index, i.e.,

$$s^0 = O(\lambda^{-1/4} T^{1/4} h^{-1/4}). \tag{22}$$

By similar considerations, one can now simplify the calculations for estimating the orders of magnitude of the local slopes at larger-scale levels ($k = 1, \ldots, l$). Practically, as will be confirmed below, in order to estimate the size of a certain c^k_i, one can introduce into (2) only the following truncation of u^0

$$u^{0,k} = \sum_{m=0}^{k} \sum_{j \in V_m} c^m_j \beta^{j,m}. \tag{23}$$

For instance, in order to examine the local amplitudes of the next level ($k=1$), one would insert $u^{0,1}$ into the Hamiltonian (2), obtaining a sum of terms like

$$\begin{aligned}&\frac{\lambda}{h^3}(c^0_{2i}-c^0_{2i-1}+\frac{1}{2}(c^1_i-c^1_{i-1}))^4+\frac{\lambda}{h^3}(c^0_{2i-1}-c^0_{2i-2}+\frac{1}{2}(c^1_i-c^1_{i-1}))^4\\&+\frac{1}{h}(c^0_{2i}-c^0_{2i-1}+\frac{1}{2}(c^1_i-c^1_{i-1}))^2+\frac{1}{h}(c^0_{2i-1}-c^0_{2i-2}+\frac{1}{2}(c^1_i-c^1_{i-1}))^2\\&=\lambda h(s^0_{2i}+s^1_i)^4+\lambda h(s^0_{2i-1}+s^1_i)^4+h(s^0_{2i}+s^1_i)^2+h(s^0_{2i-1}+s^1_i)^2.\end{aligned}\tag{24}$$

In order to estimate a certain s^1_i one would like to determine the dominant part in (24) that depends on s^1_i.

Generally, inspecting the various terms of $\mathcal{H}_h(u^{0,k})$, it turns out (see note below) that, under the assumption $h<\lambda T$, for a certain local slope s^k_i, the dominant part in $\mathcal{H}_h(u^{0,k})$ which depends on s^k_i has the following order of magnitude

$$\lambda h_k\frac{1}{2^k}\sum_{j=-2^k+1}^{0}(s^0_{2^k i+j})^2(s^k_i)^2.\tag{25}$$

Hence, probable configurations are obtained for

$$\lambda h_k(s^0)^2(s^k_i)^2=O(T).\tag{26}$$

This relation emphasizes the correlation between different scales. More precisely, s^k_i is correlated with all other scales, but it is most strongly affected by the smallest-scale fluctuations, since their slopes are largest (see (20)-(21)). Hence, by (22),

$$s^k_i=s^k=O(\lambda^{-1/4}T^{1/4}h^{1/4}h_k^{-1/2})=O(2^{-k/2}s^0)\tag{27}$$

justifying by the way our use of the truncation (23).

Of course, $\mathcal{H}_h(u^{0,k})$ includes other terms than (25), but it can be shown easily that they are less dominant than (25), therefore they would not change the order of magnitude of s^k, which is given by (27), though some of them may contribute to the *deviation* in measuring the susceptibility. Specifically, in the special case $k=l$, $\mathcal{H}_h(u^{0,l})$ contains

$$T\alpha(\mathbf{s^1_1})^2+T\varphi\mathbf{s^1_1}+T\sum_{k=1}^{l-1}\varphi_k\mathbf{s^1_1}\tag{28}$$

where

$$\alpha=\lambda T^{-1}h\sum_{i=1}^{N}(s^0_i)^2$$

$$\varphi = \lambda T^{-1} h \sum_{i=1}^{N} (s_i^0)^3$$

and

$$\varphi_k = \lambda T^{-1} h \sum_{i \in \bar{V}_k} s_i^k \sum_{j=-2^k+1}^{0} (s_{2^k i+j}^0)^2.$$

Notice that, the first term in (28) is actually (25) for the case $k = l$. Practically, since $s_{2i-1}^k = -s_{2i}^k$ $(k = 0, \ldots, l)$, φ cancels out *exactly* to zero, and φ_k is relatively small

$$\varphi_k = O(\lambda T^{-1} h^{1/2} (s^0)^2 s^k) = O(\lambda^{1/4} T^{-1/4} h^{-1/4} 2^{-k/2}).$$

Note. The following simplification was employed in the calculations mentioned above. Since the expression

$$\frac{(2h)^{-1}}{2^{k-1}} \sum_{j=2^k(i-1)+2}^{2^k i} \beta^{i,k}(x_j) c_i^k + \beta^{i-1,k}(x_j) c_{i-1}^k - \beta^{i,k}(x_{j-2}) c_i^k - \beta^{i-1,k}(x_{j-2}) c_{i-1}^k$$

(with summation running only over even integers) is exactly the local slope s_i^k, the replacement of the "original" squared slope in $\mathcal{H}_h(u^{0,k})$

$$\frac{(2h)^{-2}}{2^{k-1}} \sum_{j=2^k(i-1)+2}^{2^k i} (\beta^{i,k}(x_j) c_i^k + \beta^{i-1,k}(x_j) c_{i-1}^k - \beta^{i,k}(x_{j-2}) c_i^k - \beta^{i-1,k}(x_{j-2}) c_{i-1}^k)^2$$

by the squared slope $(s_i^k)^2$ is reasonable. Indeed, we have cofirmed empirically that this replacement would not change the orders of magnitude that were estimated above.

From (27), a reasonable estimate for c_{2i+1}^k is

$$c_{2i+1}^k = c^k = O(\lambda^{-1/4} T^{1/4} h^{1/4} h_k^{1/2}) \qquad (2i+1 \in V_k). \tag{29}$$

As in the variable-coupling Gaussian model [4], an important property of the representation (17) is the weak correlation between different local amplitudes, i.e., $\{c_i^k\}_{i \in V_k}$ are not strongly correlated to $\{c_i^n\}_{i \in V_n}$ for $k \neq n$, and far local amplitudes on the same level are weakly correlated. From this and (17), the magnetization (4) can be expressed as

$$M_h(u) = \sum_{k=0}^{l} \sum_{i \in V_k}^{*} f_i^k c_i^k$$

where $\sum^*$ denotes summation only over the odd integers and

$$f_i^k \approx f^k = \frac{h^a}{L^a} 2^k. \tag{30}$$

Hence

$$M_h^2(u) = \sum_{k=0}^{l} \sum_{n=0}^{l} \sum_{i\in V_k}^{*} \sum_{j\in V_n}^{*} f_{ij}^{kn} c_i^k c_j^n \tag{31}$$

and

$$f_{ij}^{kn} = f_i^k f_j^n \approx f^{kn} = \frac{h^{2a}}{L^{2a}} 2^{k+n}. \tag{32}$$

Clearly,

$$\langle c_i^k \rangle = 0 \tag{33}$$

and by (29)

$$\langle (c_{2i+1}^k)^2 \rangle = O(\lambda^{-1/2} T^{1/2} h^{1/2} h_k). \tag{34}$$

Using (31)-(34) one can show that

$$\begin{aligned} \langle M_h^2 \rangle &\approx \sum_{k=0}^{l} f^{kk} \sum_{i\in V_k}^{*} \langle (c_i^k)^2 \rangle \\ &\propto \frac{h^{2a}}{L^{2a}} \sum_{k=0}^{l} 2^{2k} L h_k^{-1} \lambda^{-1/2} T^{1/2} h^{1/2} h_k \\ &= \lambda^{-1/2} T^{1/2} h^{2a-2} L^{1-2a} h^{1/2} \sum_{k=0}^{l} h_k^2. \end{aligned} \tag{35}$$

Therefore, a proper scaling exponent for the discrete magnetization would be $a = 3/4$, otherwise (35) would either diverge or vanish as $h \to 0$. Thus, the discrete susceptibility (4) is defined as

$$\chi_h = \langle \frac{h^{3/2}}{L^{3/2}} (\sum_i u_i)^2 \rangle.$$

From (35) it is clear that as in the simple Gaussian models the susceptibility is indeed dominated by the smoothest component, being mainly proportional to $(c_1^l)^2$. But unlike the situation in the Gaussian models, the size of that smoothest component is affected by all other components and especially by the highest-frequency ones. This claim is quite evident from the observation in (26) for the particular case $k = l$ and $i = 1$, i.e.,

$$\lambda h_l (s^0)^2 (s_1^l)^2 = O(T).$$

4.1. Estimation of the coarsening bias in the susceptibility

Since the susceptibility is mainly proportional to $(s_1^l)^2$, it would be enough to estimate the bias in $\langle (s_1^l)^2 \rangle$. To demonstrate the bias concept, let us refer to the expression (28) as if φ is not zero and the $\varphi_k's$ are zero. Then, according to (28) it turns out that, for given values of $\{s_i^0\}$,

$$\langle (s_1^l)^q \rangle \approx \frac{\int (s_1^l)^q e^{-(\alpha (s_1^l)^2 + \varphi s_1^l)} ds_1^l}{\int e^{-(\alpha (s_1^l)^2 + \varphi s_1^l)} ds_1^l}$$

where $\alpha = \frac{\lambda}{T} \frac{1}{N} \sum_i (s_i^0)^2 = \frac{\lambda}{T} (s^0)^2$ and $\varphi = \frac{\lambda}{T} \frac{1}{N} \sum_i (s_i^0)^3 = \frac{\lambda}{T} N^{-1/2} (s^0)^3$.
Therefore, for $q = 1$ and $q = 2$,

$$\langle s_1^l \rangle^2 = \frac{\varphi^2}{4\alpha^2}$$

$$\langle (s_1^l)^2 \rangle = \frac{\varphi^2}{4\alpha^2} + \frac{1}{2\alpha}. \tag{36}$$

Clearly, using (22),

$$\langle (s_1^l)^2 \rangle = O(\lambda^{-1/2} T^{1/2} h^{1/2}).$$

Note that the quantity $\frac{\varphi^2}{4\alpha^2}$ is a *bias* in the estimation of the squared average magnetization, caused by freezing $\{s_i^0\}$; hence this quantity is also a bias in the susceptibility estimation, caused upon coarsening. It follows from (22) that the bias itself has the same order of magnitude as the observable in question, i.e.,

$$\frac{\varphi^2}{4\alpha^2} = O(\lambda^{-1/2} T^{1/2} h^{1/2}).$$

Therefore, the relative accuracy in measuring the susceptibility would not decrease forever as N grows, and it is not possible to construct an optimal algorithm for any given accuracy. Actually, at small enough grids and low accuracy levels the problem cannot be detected (as evident in our numerical experiments, see Ex. 3 in Sec. 8), since the coefficient of the bias is quite small.

In fact, as we explained before, the terms in φ *exactly* cancel each other. The problem is really introduced only at the next coarsening steps, not with terms like φ, which for convenience were discussed above, but with terms like φ_k which contribute the following bias in the squared average magnetization, caused by freezing $\{s_i^0\}$ and $\{s_i^k\}$,

$$\frac{\varphi_k^2}{4\alpha^2} = O(\lambda^{-1/2} T^{1/2} h^{1/2} 2^{-k}) \qquad k = 1, \ldots, l$$

and as a result at coarsening the susceptibility estimation is biased. Since the bias of large scale slopes is less influencing, the problem becomes apparent only at high levels of accuracy, which interfere small scale slopes.

That the bias is not really introduced at the very first coarsening step can also be seen as follow. Consider a certain site i which does not belong to the next coarser level ($i = 2j + 1$). The anharmonic part of the Hamiltonian in the neighborhood of that site is proportional to

$$\begin{aligned}
&(u_i - u_{i-1})^4 + (u_{i+1} - u_i)^4 \\
&= \frac{1}{8}(2u_i - u_{i-1} - u_{i+1})^4 \\
&+ \frac{1}{8}(u_{i+1} - u_{i-1})^4 + \frac{3}{4}(2u_i - u_{i-1} - u_{i+1})^2(u_{i+1} - u_{i-1})^2.
\end{aligned} \tag{37}$$

Since the interpolation operator I_1^0 is the equal-weight linear operator (.5,.5), the quantity $2u_i - u_{i-1} - u_{i+1}$ is frozen at coarsening, hence the coarse grid Hamiltonian, given by terms like (37), still retains the sign symmetry, under which $\langle s_1^l \rangle = 0$. Indeed, two-level tests confirm that the first coarsening is not disturbing the feasibility of an optimal algorithm (see Ex. 6 in Sec. 9). Our main conclusion at this stage: in order to devise an optimal algorithm, we have selected the *most effective linear* interpolation operators (see Sec. 3.3), but it turns out that these linear operators are not effective enough. It is thus natural to go one step further by constructing *nonlinear* interpolation operators (see Sec. 7), which indeed will prove much more efficient (see Sec. 8).

5. Optimal choice of the multigrid cycling parameters

As mentioned, the susceptibility is dominated by the largest-scale fluctuations, but unlike the simple Gaussian case, the size of these fluctuations is affected by other components, especially by the smallest-scale components.

We would like to develop a multigrid algorithm that will efficiently sample the susceptibility. It is shown here that the optimal cycle index is not fixed through the different levels, i.e., the cycle indices may change from one current level to another and moreover it is not necessarily an integer number. The way to implement this unusual multigrid cycle is given in Sec. 3.4 above (see also Ref. 3).

We next study the number m_k of relaxation sweeps the algorithm needs to perform on level k, i.e., on a grid with meshsize $h_k = 2^k h$, ($k = 0, \ldots, l = \log_2(N/2)$), in order to achieve relative accuracy $\varepsilon_r = \varepsilon/\sigma$ in the estimation of the susceptibility. We first ignore the inter-scale dependence, described above, in which case we will get a bound on the cycle indices γ_c that are needed at coarser levels of the algorithm. Clearly, the number m_k depends on the contribution of the coefficients $\{c_i^k\}$ (amplitudes of components with scale length of order h_k) to the deviations in measuring the susceptibility, because, as indicated in Sec. 4, only a relaxation sweep on level k samples those coefficients efficiently.

Consider a general term (c_i^k, c_j^n) in (31). Using the weak correlation between distinct stochastic variables, with (30), (32), (33) and (34) above, the standard

deviation of such a term from its average is

$$\sigma(f_{ij}^{kn}c_i^k c_j^n) \approx f_{ij}^{kn}\sigma(c_i^k)\sigma(c_j^n) = O(\lambda^{-1/2}T^{1/2}L^{-3/2}h_k^{3/2}h_n^{3/2}).$$

For the case, $h_k = h_n$, the term (c_i^k, c_j^k) in (31) is effectively sampled m_k times in a cycle. There are $O(h_k^{-2}L^2)$ such terms, which are almost uncorrelated, hence their total contribution to the deviation is

$$O(m_k^{-1/2}\lambda^{-1/2}T^{1/2}L^{-1/2}h_k^2) \qquad (k = 0, \ldots, l). \tag{38}$$

For the case $h_k > h_n = h_{k-r}$ $(r \geq 1)$ with $f^k\sigma(c_i^k) > f^n\sigma(c_j^n)$ (see (30) and (34)), the term (c_i^k, c_j^n) is then sampled at least m_k times in a cycle. Therefore, and since for a given k and n with $f^k\sigma(c_i^k) > f^n\sigma(c_j^n)$ there are $(h_k^{-1}h_n^{-1}L^2)$ such terms which are almost uncorrelated, hence their total deviation in a cycle is bounded by

$$O(m_k^{-1/2}\lambda^{-1/2}T^{1/2}L^{-1/2}h_k h_{k-r}) \qquad (r \geq 1). \tag{39}$$

For a given level k, summation of the contributions to the deviation (39) over integers $r \geq 0$, gives again the estimation in (38). Hence, the total *relative* expected error in measuring the susceptibility is

$$\varepsilon_r = \varepsilon/\sigma = O(\varepsilon/\langle M_h^2\rangle) = O(\sum_{k=0}^{l} m_k^{-1/2}h_k^2) \tag{40}$$

while the total work (operations) on all level is clearly

$$W = O(\sum_{k=0}^{l} m_k L h_k^{-1}). \tag{41}$$

The optimal choice for m_k (yielding either minimal ε_r for a given W or minimal W for a given ε_r) is obtained when $\frac{\partial\varepsilon_r}{\partial m_k} + \lambda_1\frac{\partial W}{\partial m_k} = 0$, which by (40) and (41), yields

$$m_k = \lambda_2 h_k^2 = \lambda_3 2^{2k}, \tag{42}$$

where λ_1,λ_2 and λ_3 are independent of k. Relation (42) is realized by the optimal cycle index $\gamma_c = 4$.

For *any* fixed cycle index γ_c we have $m_k = m\gamma_c^k$, where m is the total number of cycles performed. Since $h_k = L2^{k-l-1}$, we can perform the summations in (40) and (41) and obtain

$$\varepsilon_r = O\left(m^{-1/2}\frac{\gamma_c^{-l/2} - 2^{-2(l+1)}\gamma_c^{1/2}}{1 - 2^{-2}\gamma_c^{1/2}}L^2\right)$$

and

$$W = O\left(m\frac{\gamma_c^l - 2^{l+1}\gamma_c^{-1}}{1-2\gamma_c^{-1}}\right)$$

provided $2 < \gamma_c < 16$. Thus, any $2 < \gamma_c < 16$, not just $\gamma_c = 4$, yields $W = O(\varepsilon_r^{-2})$. Asymptotically (for $\varepsilon_r \to 0$), the minimal value of $W\varepsilon_r^2$ is attained for $\gamma_c = 4$ and values very close to the minimum are obtained for cycle indices which are close to 4. Since $\gamma_c > 2$ the work is dominated by the coarsest grid work, implying

$$m_l = O(\varepsilon_r^{-2}). \tag{43}$$

Observe, though, that if γ_c is considerably larger than 2 for, say, half the levels (e.g., the coarse ones) it can be considerably less than 2 for the other half and still optimality, in the sense of $W = O(\varepsilon_r^{-2})$ is attained.

So far, however, we have neglected in this analysis the influence of high-frequency and mid-frequency components on the amplitudes of the smooth ones. In particular, consider the dominant amplitude of s_1^l, which is given by (28), i.e., the stochastic amplitude of s_1^l from all levels is expressed in the Hamiltonian by

$$\alpha(s_1^l)^2 + \tilde{\varphi}s_1^l$$

where

$$\tilde{\varphi} = \sum_{k=1}^{l-1}\varphi_k.$$

Therefore, the dependence of the dominant amplitude of s_1^l, on the smallest-scale and the mid-scale fluctuations is given by

$$\langle (s_1^l)^2\rangle = \frac{\tilde{\varphi}^2}{4\alpha^2} + \frac{1}{2\alpha}. \tag{44}$$

Since $\frac{\varphi_k^2}{4\alpha^2}$ is small relatively to $\frac{1}{2\alpha}$, we would like to neglect the first term in (44), for a while, then the standard deviation of this stochastic amplitude (α) over the m_0 samples on the finest grid is

$$O(\frac{\lambda}{T}m_0^{-1/2}N^{-1/2}(s^0)^2).$$

Therefore, it is necessary that the relative accuracy of that level would satisfy

$$O(m_0^{-1/2}N^{-1/2}) \le \varepsilon_r. \tag{45}$$

That would mean

$$m_0 N \ge O(\varepsilon_r^{-2}).$$

Since we want our overall work to be at most $O(\varepsilon_r^{-2})$, we must have

$$m_0 N = O(\varepsilon_r^{-2}). \tag{46}$$

With (43) it is clear that, the work on the coarsest grid should be proportional to the work on the finest grid.

Next, we would like to consider $\tilde{\varphi}$. Since α is frozen at coarsening it turns out that in order to determine m_k at any mid-level k, it would be enough to examine the standard deviation of

$$(\tilde{\varphi})^2 = \sum_{i=1}^{l-1}\sum_{j=1}^{l-1}\varphi_i\varphi_j. \tag{47}$$

Recall that

$$\varphi_k = \lambda T^{-1} h \sum_{i\in\bar{V}_k} s_i^k \sum_{j=-2^k+1}^{0} (s^0_{2^k i+j})^2 = O(\lambda T^{-1} h^{1/2}(s^0)^2 s^k). \tag{48}$$

Consider a general term (φ_i, φ_j) in (47). Since φ_i and φ_j are weakly correlated for $i \neq j$ and $\langle\varphi_k\rangle = 0$, the standard deviation of such a term from its average is

$$\sigma(\varphi_i\varphi_j) = \sigma(\varphi_i)\sigma(\varphi_j) = \langle\varphi_i^2\rangle^{1/2}\langle\varphi_j^2\rangle^{1/2} = O(T^{-2}\lambda^2 h(s^0)^4 s^i s^j)$$

(note that the last estimation holds also for the $i = j$ case).

A term (φ_i, φ_i) is effectively sampled m_i times in a cycle. For the case $j > i$ with $\sigma(\varphi_i) > \sigma(\varphi_j)$ (see (48)), the term (φ_i, φ_j) is sampled at least m_i times in a cycle. Therefore, the total expected error in measuring $(\tilde{\varphi})^2$ is

$$\begin{aligned}
\varepsilon &= O\Big(\sum_{i=1}^{l-1}\sum_{j\geq i} m_i^{-1/2}\langle\varphi_i^2\rangle^{1/2}\langle\varphi_j^2\rangle^{1/2}\Big) \\
&= O\Big(T^{-2}\lambda^2 h(s^0)^4 \sum_{i=1}^{l-1}\sum_{j\geq i} m_i^{-1/2} s^i s^j\Big) \\
&= O\Big(T^{-2}\lambda^2 h(s^0)^4 \sum_{i=1}^{l-1} m_i^{-1/2}(s^i)^2\Big).
\end{aligned}$$

Thus, the total *relative* expected error in measuring $\tilde{\varphi}^2$ is

$$\varepsilon_r = O(\varepsilon/\tilde{\varphi}^2) = \varepsilon/O(T^{-2}\lambda^2 h(s^0)^4(s^1)^2) = O(\sum_{i=1}^{l-1} m_i^{-1/2}\frac{(s^i)^2}{(s^1)^2}) = O(\sum_{i=1}^{l-1} m_i^{-1/2} 2^{-i}). \tag{49}$$

The optimal choice for m_k (yielding minimal W for a given ε_r) is obtained when $\frac{\partial \varepsilon_r}{\partial m_k} + \lambda_1 \frac{\partial W}{\partial m_k} = 0$, which by (49) and (41), yields optimal cycle index $\gamma = 1$. Although we have obtained an optimal value for γ, which means that for a given relative accuracy the least work would be done, this will not guarantee a statistically optimal algorithm, i.e., $W = O(\varepsilon_r^{-2})$. Indeed, as explained in Sec. 4.1, a coarsening bias in measuring the susceptibility is introduced by $(\tilde{\varphi})^2$, and as a result this algorithm will not achieve optimal performance. The results, presented in Sec. 8, indeed start to deteriorate at large lattices.

This observation can also be seen as follows. To ensure that the error contributed by any level $k+1 \geq 2$ is smaller than that of the next finer level k

$$\frac{m_{k+1}^{-1/2}(s^{k+1})^2}{m_k^{-1/2}(s^k)^2} < 1.$$

The last demand suggests to apply cycle index $\gamma_f > 1/4$ on fine levels. Clearly, γ_f should be smaller than 2 to keep the work on the finest level dominant. Thus, at the fine levels one should apply,

$$1/4 < \gamma_f < 2.$$

Moreover, to assure that the deviations from the different levels do not accumulate unboundedly, one should require that the relative error on level 1 is smaller than the relative error on the finest level ($k = 0$), in other words,

$$O\left(\frac{m_1^{-1/2}}{m_0^{-1/2} N^{-1/2}}\right) < 1$$

or

$$O\left(\frac{N}{m_1/m_0}\right) < 1. \tag{50}$$

Clearly, a bounded cycle index can not satisfy the last condition for all N. This explains why statistically optimal efficiency is not achieved for any grid size.

Since the relative error on the finest level is of order ε_r (45), then the total relative error in measuring the stochastic amplitude is comparable to the relative accuracy in measuring the susceptibility, i.e., $O(\varepsilon_r)$, in cases when (50) is satisfied.

In our numerical experiments (see Ex. 3 in Sec. 8) $l/2$ fine grids would apply γ_f and the rest $l/2$ coarse grids would apply

$$\gamma_c = 4/\gamma_f$$

satisfying $2 < \gamma_c < 16$ and γ_c is close to 4. As a result, the work on the coarsest grid is proportional to the work on the finest grid and the geomtric mean of the cycle indices is approximately 2. Moreover, the values of γ_c and γ_f and the relations in (43) and (46) would enforce the optimal relation

$$W = O(\varepsilon_r^{-2})$$

only in cases that satisfy (50).

6. The unigrid algorithm with nonlinear minimization interpolation operator

For relative high accuracy, both the analysis and the numerical tests show that ideal performance can no longer be obtained by a multigrid process which employs our weighted linear interpolation. At this stage, we have introduced a novel nonlinear interpolation, called *minimization interpolation*, which is very effective. The minimization interpolation is very efficient in the sense that a stochastic coarse grid change is followed by an appropriate most probable change on the finer grid. The most probable fine grid change is the one that minimizes the local fine grid Hamiltonian for the *given* coarse grid change. Here, for start, the concept is given only for two levels, for which we easily applied a *unigrid* technique. In Sec. 7, our complete *multigrid* algorithm that employs the nonlinear minimization interpolation operator is described in details.

6.1. Definition of the nonlinear minimization interpolation operator

The general idea of the minimization interpolation, for any model at any dimension, is as follows: in order to define a value v_0 at a fine-grid point based on coarse grid values $(v_1, v_2, \ldots)$, the minimization interpolation method is first to calculate $V_0(v_1, v_2, \ldots)$, defined as the value of v_0 that would be obtained by some, exact or approximate, local Hamiltonian minimization with the value of $(v_1, v_2, \ldots)$ being held fixed. Then, to retain statistical detailed balance, the minimization interpolation value is defined by

$$v_0 = V_0(v_1, v_2, \ldots) + \tilde{v}_0 - V_0(\tilde{v}_1, \tilde{v}_2, \ldots)$$

where the $\tilde{v}_i$ are the values of the variables at coarsening.

Consider an anharmonic Hamiltonian with *non-uniform* positive couplings, i.e.,

$$H(u) = \sum_{i=1}^{N} f_i (u_i - u_{i-1})^4 + \sum_{i=1}^{N} g_i (u_i - u_{i-1})^2 .$$

Actually, this model simulates the coarse-grid Hamiltonian after the first coarsening, when I_1^0 is the pure linear interpolation operator. In this particular case, the f_i couplings of the coarse-grid Hamiltonian are the same for the whole interval, and the g_i couplings are varying over the interval.

The fine grid configuration at transition to the next coarser level will be denoted by $\{\tilde{u}_i\}_{i=0}^{N}$. The next coarser level is constructed by taking every other configuration spin, i.e., $\{u_{2i}\}_{i=0}^{N/2}$. The question is how to interpolate effectively a change from the next coarser level to the original grid, or in other words how to specify effective values to $\{u_{2i+1}\}_{i=0}^{i=N/2-1}$ when $\{u_{2i}\}_{i=0}^{N/2}$ are given.

Let β denotes the following known quantity

$$\beta = u_{2i+2} - u_{2i}$$

then we seek for the *unknown*

$$\alpha^* = u^*_{2i+1} - u_{2i}$$

that would minimize the local energy.

In other words, α^* would satisfy

$$E_\alpha(\alpha^*) = 0 \tag{51}$$

where

$$E(\alpha) = f_{2i+1}\alpha^4 + g_{2i+1}\alpha^2 + f_{2i+2}(\beta-\alpha)^4 + g_{2i+2}(\beta-\alpha)^2. \tag{52}$$

Then, in order to preserve symmetry between the two configurations $\{\tilde{u}_i\}_{i=0}^N$ and $\{u_i\}_{i=0}^N$, u_{2i+1} is computed as follows

$$u_{2i+1} = u^*_{2i+1} + s \tag{53}$$

where s is the distance or the shift of $\tilde{u}_{2i+1}$ from the value which minimizes the local energy before transition to the coarse grid. (Determine the parameter $\tilde{\beta} = \tilde{u}_{2i+2} - \tilde{u}_{2i}$ and to obtain $\tilde{\alpha}^* = \tilde{u}^*_{2i+1} - \tilde{u}_{2i}$ solve $\tilde{E}_{\tilde{\alpha}}(\tilde{\alpha}^*) = 0$, where $\tilde{E}(\tilde{\alpha}) = f_{2i+1}\tilde{\alpha}^4 + g_{2i+1}\tilde{\alpha}^2 + f_{2i+2}(\tilde{\beta}-\tilde{\alpha})^4 + g_{2i+2}(\tilde{\beta}-\tilde{\alpha})^2$. Then, s would be $\tilde{u}_{2i+1} - \tilde{u}^*_{2i+1}$.)

Two-level unigrid tests are presented in Sec. 9 (Ex. 10).

Note two important observations. First, the local energy (52) has one and only one minimum for a given β. Second, in the case that the couplings satisfy $f_{2i+1} = f_{2i+2}$ and $g_{2i+1} = g_{2i+2}$ for $i = 0, ..., N/2-1$, the nonlinear interpolation operator coincides with the pure linear interpolation.

6.2. Metropolis method

The relaxation, on both levels, has been done by the Metropolis rule [23]. On the original grid, a relaxation step at site i is done as follows. Let u^{old} be the previous configuration and u^{new} the candidate configuration obtained by changing u_i^{old} to u_i^{new}, which is selected randomly uniformly in the interval $[u_i^{old} - \delta, u_i^{old} + \delta]$, where δ is a parameter. Then, changing is decided, with the following "transition probability":

$$P(u^{old} \to u^{new}) = \min\left[\frac{P(u^{new})}{P(u^{old})}, 1\right]. \tag{54}$$

On the coarse level, the metropolis relaxation is of course applied only at even sites. One has to take into account that a relaxation step at spin u_{2i}, leads also to changes at u_{2i-1} and u_{2i+1}, as explained in the previous section.

The size of the parameter δ is not the same for both levels. Its size is determined as a function of the acceptance rate in each level; the parameter δ is chosen such that the acceptance rate is about 50% (the acceptance rate is the ratio between the number of acceptances of the candidate u^{new} and the total number of random trials).

An important issue is the relative weights that should be associated with coarse level configuration densities. We decompose a fine-grid configuration t to coarse-grid variables and fine-grid variables *only*, i.e., a configuration t

$$t = (t_0, t_1, \ldots, t_N)$$

is expressed by

$$\bar{t} = (u, v) = (u_0, u_1, \ldots, u_{N/2}, v_1, \ldots, v_{N/2})$$

where $u_i = t_{2i}$ and $v_i = t_{2i-1}$.

Recall that on the *fine-grid* for given values

$$\xi = (\xi_0, \ldots, \xi_{N/2}) \quad d\xi = (d\xi_0, \ldots, d\xi_{N/2})$$

$$\eta = (\eta_1, \ldots, \eta_{N/2}) \quad d\eta = (d\eta_1, \ldots, d\eta_{N/2})$$

where $\| d\xi \|$ and $\| d\eta \|$ are infinitesimally small, the probability that (u, v) would satisfy

$\xi_i \le u_i \le \xi_i + d\xi_i \quad i = 0, \ldots, N/2$ and $\eta_j \le v_j \le \eta_j + d\eta_j \quad j = 1, \ldots, N/2$

is given directly by the Boltzmann density function, i.e.,

$$p\{\xi_i \le u_i \le \xi_i + d\xi_i, \eta_j \le v_j \le \eta_j + d\eta_j\} = P(\xi, \eta) d\xi d\eta.$$

Now, assume that we can have on the coarse grid an explicit way to express v_j as a function of $\xi_j - \xi_{j-1}$ and the shift s_j, e.g.,

$$v_j = F_j(\xi_j - \xi_{j-1}, s_j).$$

Therefore, on the *coarse-grid*

$$\begin{aligned}
&p\{\xi_i \le u_i \le \xi_i + d\xi_i, F_j(\xi_j - \xi_{j-1}, s_j) \le v_j \le F_j(\xi_j - \xi_{j-1}, s_j + ds_j)\} \\
=&p\{\xi_i \le u_i \le \xi_i + d\xi_i, F_j(\xi_j - \xi_{j-1}, s_j) \le v_j \le F_j(\xi_j - \xi_{j-1}, s_j) + \frac{\partial F_j(\xi_j - \xi_{j-1}, s_j)}{\partial s_j} ds_j\} \\
=&P(\xi, F(\bigtriangledown \xi, s)) \Pi_i \frac{\partial F_i(\xi_i - \xi_{i-1}, s_i)}{\partial s_i} d\xi ds.
\end{aligned}$$

Clearly, from the definition of the minimization interpolation, F_i is linear with respect to s_i and

$$\frac{\partial F_i(\xi_i - \xi_{i-1}, s_i)}{\partial s_i} = 1.$$

Therefore, the probability density function on the coarse grid is

$$P(\xi, F(\bigtriangledown\xi, s))$$

without any weight factor multiplication.

This derivation shows that the transition probability (54) is exactly the ratio between the physical densities of the candidate configuration and the previous configuration when applied either on the fine level or on the coarse level. This observation is used below to prove the detailed balance condition.

6.3. Proof of detailed balance condition

It is sufficient to prove that the stochastic process on the coarser level preserves symmetry between any two configurations. In other words, we show that if B is a candidate fine-level configuration obtained from another fine-level configuration A in the coarse-level relaxation step at site $2i$, then there is the same chance to get A as a candidate obtained from B in that relaxation step, even if the at-coarsening variables in the two cases were different. From the definition of the Metropolis rule, it is clear that if B_{2i} is obtained uniformly from the interval $[A_{2i} - \delta, A_{2i} + \delta]$ with density $\frac{1}{2\delta}$, then A_{2i} would be obtained with the same density from $[B_{2i} - \delta, B_{2i} + \delta]$. By definition of the minimization interpolation, the shift s in B is the same as the shift in A (53), and by the observation that the local energy (52) has one and only one minimum for given coarse-grid values, it is clear that if the candidate spin at site $2i$ is A_{2i} then its two neighbors would be A_{2i-1} and A_{2i+1}. With the observation above that the transition probability (54) is exactly the ratio between the physical probabilities, independently of the at-coarsening variables, it is clear that the detailed balance condition is satisfied.

7. Devising a multigrid algorithm with nonlinear minimization interpolation operators

Our main aim is to construct an optimal algorithm, in the sense that a relative accuracy ε_r is obtained in $O(\varepsilon_r^{-2})$ operations. As explained and demonstrated before, it is essential to use an effective interpolation operator, such as the nonlinear minimization interpolation operator. The two-level unigrid experiment, Ex. 10 in Sec. 8, implies the feasibility to construct an optimal algorithm. But to employ a unigrid cycle would be of course very expensive, especially due to the fact that for statistical purposes one would better do many more sweeps on coarse scales than on fine scales. To overcome this complexity disadvantage it is required to derive an

explicit coarse level Hamiltonian, which can be complicated. Since the minimization interpolation operator is given by a nonlinear implicit scheme, it would be impossible to generate exact coarse level Hamiltonians. Therefore, using elements from approxiamtion theory [10], coarse level Hamiltonians are approximated by polynomials. Generally, for smooth functions, arbitrarily close approximations can be attained by raising the polynomial order.

To facitilate a recursive multigrid cycle, the Hamiltonians on all levels should better have the same general structure. Therefore, at first one has to decide what is the order of the approximating polynomial (e.g., $p = 2, 4, 6$, etc.) and then to adjust the relaxation module, the interpolation module and the coarsening module. The multigrid algorithm, which will be described below is slightly different from the ultimate optimal algorithm, since we did not construct an estimator for the sucseptibility. Such an estimator should be defined in such a way that the susceptibility, too, can be evaluated on any level, without going back to finer levels. In our current multigrid algorithm, statistics measurements are done only through transfers to the finest level. This means that a relaxation on a certain level k should be followed by the operation $m_1^0 m_2^1 \ldots m_k^{k-1}$, where m_i^{i-1} is the minimization interpolation from the coarse level i to the next finer level $i-1$.

7.1. The relaxation module

Here, on all levels ($k = 0, \ldots, l = \log_2(N/2)$), the relaxation is done according to the Metropolis rule (see Sec. 6.2). A transition from a fine level k to the next coarser level $k+1$ defines a coarse grid Hamiltonian which is fixed through the processing on the coarse level. Thus, a long Monte Carlo process can be done on the coarse level, without going back to finer levels. The coarse grid Hamiltonian is a FAS Hamiltonian (see Sec. 7.2 below), and the coarse grid variables are initially set to the corresponding current fine grid values. Since the Hamiltonian at all levels has the same general structure (p—order polynomial), the Monte Carlo module is the same for all levels. Its steps on level k in the multigrid algorithm involve only the k^{th} grid Hamiltonian, so only values on the current coarse grid are needed to decide whether to update a certain variable.

As in Sec. 6.2, the range parameters of the Metropolis rule must be specified. In each level i, δ_i is chosen according to the 50% acceptance rate criteria at this level. Practically, having the range parameters for small lattice, e.g. $N = 16$, one can obtain reasonable values for the range parameters on a lattice twice larger, i.e., $N = 32$, by using the estimate in (29). Then, a refinement of the range parameters can be done. Having the tuned parameters for the larger lattice, one can continue in the same manner for larger lattices.

7.2. The minimization interpolation module

This interpolation is best defined in terms of the Full Approximation Scheme (FAS; cf Sec. 7.1 in Ref. 6), where the coarse-grid variables represent the full current configuration instead of just the coarse-grid displacement (as in Sec. 3.1).

The definition of the nonlinear minimization interpolation from level k to level $k-1$ is generalized for every $p-$polynomial Hamiltonian which is associated with the fine level $(k-1)$, as follows.

Consider a $p-$polynomial Hamiltonian on level $k-1$, i.e.,

$$H(u) = \sum_{i=1}^{N} \sum_{j=0}^{p} a_i^j (u_i - u_{i-1})^j .$$

(Except for the couplings of the leading terms, the couplings are not necessarily positive.)

Let us denote the coarse-grid values on level k by $\{U_I\}_{I=0}^{N/2}$, and the at-coarsening fine-grid values on level $k-1$ by $\{\tilde{u}_i\}_{i=0}^{N}$. Then, the fine-grid configuration after interpolation updating $\{u_i\}_{i=0}^{N}$ is obtained as follows. First, the values of the even fine-grid variables $\{u_{2I}\}_{I=0}^{N/2}$ are simply $\{U_I\}_{I=0}^{N/2}$, respectively. The values of the odd fine-grid variables $\{u_{2I+1}\}_{I=0}^{N/2-1}$ are obtained in the following manner. Let β denotes the folllowing known quantity

$$\beta = U_{I+1} - U_I = u_{2I+2} - u_{2I}$$

then we seek the unknown

$$\alpha^* = u^*_{2I+1} - U_I = u^*_{2I+1} - u_{2I}$$

that would minimize the local energy, i.e.,

$$E_\alpha(\alpha^*) = 0 \tag{55}$$

where

$$E(\alpha) = \sum_{j=0}^{p} a^j_{2I+1} \alpha^j + \sum_{j=0}^{p} a^j_{2I+2} (\beta - \alpha)^j . \tag{56}$$

Then, under the assumption that the local energy (56) has one and only one minimum for a given β, in order to preserve detailed-balance, u_{2I+1} is computed as follows

$$u_{2I+1} = u^*_{2I+1} + s$$

where s is the distance of $\tilde{u}_{2I+1}$ from the value which minimizes the local energy at coarsening (as in Sec. 6.1). Note that in cases where the minimum is not unique, one can introduce a criteria in order to choose α^*, e.g., α^* would be the smallest α for which the minimum of $E(\alpha)$ is obtained.

The solution α^* is a root of $(p-1)-$order polynomial, which can easily be computed directly; e.g., by few steps of Newton or the secant method, starting from its former value.

Later on, the function notation f_s will be used for the minimization interpolation function, i.e.,

$$u_{2I+1} - u_{2I} = f_s(u_{2I+2} - u_{2I}) = \alpha^* + s. \tag{57}$$

7.3. The coarsening module - computation of approximate Hamiltonians on coarse levels

Since the finest-grid Hamiltonian has uniform couplings (2), it turns out that the minimization interpolation from level 1 to level 0 is a pure linear interpolation operator. Therefore, the Hamiltonian on level 1 can be expressed directly in its exact form (10), which is again a 4^{th} order polynomial. This means that the approximate Hamiltonians need be generated only for levels $k = 2, \ldots, l = \log_2(N/2)$.

We calculate the approximate Hamiltonians according to the theory of functional approximation [10], as follows. We concentrate on a typical triplet of neighboring spins $u_{2i}, u_{2i+1}, u_{2i+2}$, which will be called without loss of generality ϕ_0, ϕ_1, ϕ_2, in order to explain how to generate an approximate local coarse grid Hamiltonian, in terms of ϕ_0 and ϕ_2 only.

Generally, the local action on the fine grid k is given by

$$E(\phi_1 - \phi_0, \phi_2 - \phi_1) = \sum_{i=0}^{p} b_i(\phi_1 - \phi_0)^i + \sum_{i=0}^{p} c_i(\phi_2 - \phi_1)^i$$

and we like to compute coefficients $a_0, a_1, \ldots, a_p$ s.t.

$$V(\phi_2 - \phi_0) = E(f_s(\phi_2 - \phi_0), \phi_2 - \phi_0 - f_s(\phi_2 - \phi_0)) \approx \sum_{i=0}^{p} a_i(\phi_2 - \phi_0)^i$$

with small errors especially in the range of probable values of $\phi_2 - \phi_0$, where f_s is the minimization interpolation function (57). This means that the $p-$order polynomial approximation will be a very good approximation to the local energy $V(\phi_2 - \phi_0)$ on level $(k+1)$ in a finite interval $\phi_2 - \phi_0 \in [x_0, x_m]$, say, over which $V(\phi_2 - \phi_0)$ yields relatively probable density, i.e., $e^{-V/T} = O(e^{-V_{\min}/T})$. Outside this interval it is enough to require that the approximation will yield large values (compared with the values inside the interval), no matter how close they are to the real local energy $V(\phi_2 - \phi_0)$.

The construction of

$$T(\phi_2 - \phi_0) = \sum_{i=0}^{p} a_i(\phi_2 - \phi_0)^i$$

can be described by the following 3 stages:

1. Calculate for $i = 0, \ldots, m$ the quantity $V_i = V(x_i) = E(f_s(x_i), x_i - f_s(x_i))$, where $\{x_i = (\phi_2 - \phi_0)_i\}_{i=0}^m$ is a finite set of $m+1$ values in the interval $[x_0, x_m]$. Note, that for a certain local triplet, the shift s is the same for any value of $x_i = (\phi_2 - \phi_0)_i$, since s depends on the at-coarsening fine-grid configuration.

2. Calculate the correspondings weights $w_0, w_1, \ldots, w_m$, where the weight function is proportional to the physical density

$$w(\phi_2 - \phi_0) = e^{-V(\phi_2 - \phi_0)/T}.$$

3. Minimize the *weighted* mean-square error Q (where $h_j = x_j - x_{j-1}$)

$$Q = \sum_{j=1}^{m} w_j (V_j - \sum_{i=0}^{p} a_i x_j^i)^2 h_j$$

by solving the linear equations

$$\frac{\partial Q}{\partial a_i} = 0 \qquad i = 0, \ldots, p$$

for the unknowns $a_0, a_1, \ldots, a_p$.

We thus construct a p–order polynomial approximation which is the best L_2 weighted approximation to the local energy $V(\phi_2 - \phi_0)$. Clearly, it is also important to keep the outer interval $(-\infty, \infty) - [x_0, x_m]$ improbable. Under the assumption that V (and hence T) in the probable area contains one and only one minimum, we have chosen the following criterion. First, check whether $T(\phi_2 - \phi_0)$ has extremum points in the outer interval. If it does, we construct a new approximating polynomial while considering some points in the outer interval with relatively small weights, to force the new approximating polynomial to be improbable in the outer area.

We have implemented the case $p = 4$ (see Ex. 5 in Sec. 8) and indeed optimal results have been achieved. The approximation calculation has used $m = 20$ (independently of the level number k) and $h_j \equiv h = \frac{x_m - x_0}{m}$ (note that x_0 and x_m do depend on k). The 5×5 linear system was solved directly. For this specific case, it is straightforward to check whether $T(\phi_2 - \phi_0)$ has any extremum point in the outer interval, since it involves just checking whether the number of real roots of the cubic equation $\frac{dT}{d(\phi_2 - \phi_0)} = 0$, is more than one. If $T(\phi_2 - \phi_0)$ has extremum points in the outer interval, then in order to construct a new approximating polynomial that keeps the outer interval non-probable, we do the following. A few number of points (≈ 10) from each side of the probable interval are added, at distance h from each other, and the number of samples in the probable interval is multiplied by a factor about 5. Then, the weighted mean-square error

Q is constructed again, and so the Euler minimization equations. To ensure that the outer points are not introduced with too high weights, we took the weight for any outer point to be the smallest between 10^{-8} and the physical weight given in stage 2. If after the reconstruction, the criterion of having only one minimum is still not satisfied, then we add more outer points and so on, till we have an approximate polynomial which satisfies the condition. Experimentally, it is shown that the number of trials is bounded independently of N. These ad-hoc procedures for approximating $V(\phi_2 - \phi_0)$ can no doubt be replaced by simpler and shorter ones.

8. Multilevel numerical results

We have applied various kinds of multigrid algorithms, with simple cycle indices or compound cycle indices, with weighted linear interpolation or nonlinear minimization interpolation, with exact Hamiltonians or approximate ones. The experiments treat the anharmonic crystal model with dominating anharmonic oscillator, except for Ex. 1 which treats a case where the harmonic oscillator is dominating. All the experiments have used temperature $T = 1$.

Three kinds of observables have been measured:

the average energy per degree of freedom

$$E_h = \langle \frac{1}{N-1} (\frac{\lambda}{h^3} \sum (u_i - u_{i-1})^4 + \frac{1}{h} \sum (u_i - u_{i-1})^2) \rangle$$

the harmonic susceptibility

$$\langle \frac{1}{N^2} (\sum u_i)^2 \rangle$$

and the anharmonic susceptibility

$$\chi_h = \langle \frac{1}{N^{3/2}} (\sum u_i)^2 \rangle$$

where each of them is associated with a corresponding thermodynamic limit.

The numerical experiments, given below, serve as a tool for demonstrating the application of the algorithms that we have developed for the anharmonic crystal model. The observables have been measured over one cycle, except for Ex. 3 and 4, where the anharmonic susceptibility was measured over two compound multigrid cycles. In order to test the optimality of a certain application, we measure the *relative* accuracy $\varepsilon_r = \frac{|\bar{A}_h - \langle A_h \rangle|}{\sigma_h(A)}$, where A_h is an observable on a (finest) grid with meshsize h, $\bar{A}_h$ is the average of its many measurements within one multigrid cycle (or two cycles), $\langle A_h \rangle$ is a long run average of the observable, and $\sigma_h(A)$ denotes its standard deviation. We say that optimality is achieved when the use of a finest grid twice finer (N twice larger), while increasing the overall work by a factor R (e.g, for a multigrid cycle with fixed cycle index $\gamma > 2$, $R = \gamma$), yields a relative error which is smaller by a factor of $\sqrt{R}$; so that relative accuracy ε_r is obtained in $O(\varepsilon_r^{-2})$ operations, independently of the grid size. In Ex. $1 - 4$ ε_r is averaged over an ensemble of 4000 runs. In Ex. 5 ε_r is averaged over an ensemble of 1000 runs only.

Experiment 1.

The relative accuracy ε_r in measuring the harmonic susceptibility with cycle index $\gamma = 3$ using the weighted linear interpolation, for $\lambda = 0.001 \cdot h$

N	ε_r
4	0.4360
8	0.2706
16	0.1561
32	0.09395
64	0.05360
128	0.03041
256	0.01796
512	0.01040
1024	0.006021
2048	0.003461
4096	0.002012

In this case the harmonic oscillator is dominating, therefore applying a multigrid cycle with cycle index greater than 2 yields statistically optimal results, as in the simple Gaussian case [3,5], i.e., the ratio between succesive ε_r entries in the table tends to $\sqrt{3} = \gamma^{1/2}$ as N becomes larger.

Experiment 2.

The relative accuracy ε_r in measuring the energy per degree of freedom,
with cycle index $\gamma = 1$ using the weighted linear interpolation,
for $\lambda = 1$

N	ε_r
4	1.0470
8	0.6603
16	0.4542
32	0.3342
64	0.2386
128	0.1709
256	0.1217
512	0.08674
1024	0.06065
2048	0.04401
4096	0.03076
8192	0.02152
16384	0.01541
32768	0.01088

Here, the anharmonic oscillator is dominating, the ratio is approximately $\sqrt{2}$. This experiment demonstrates that optimal behaviour can be achieved while using the weighted linear interpolation, when the observable is heavily dominated by the high oscillatory components.

Experiment 3.

The relative accuracy ε_r in measuring the anharmonic susceptibility, with $\gamma_f = 1.5$, $\gamma_c = 2.666$ using the weighted linear interpolation, for $\lambda = 1$

N	ε_r
8	0.2791
32	0.1720
128	0.09367
512	0.05106
2048	0.03142
8192	0.02292
32768	0.02029

This experiment justifies our claim, that the anharmonic susceptibility can be measured in an optimal time (the ratio is roughly 2) when N is relatively small, while using a compound multigrid cycle with the weighted linear interpolation. But, when N becomes larger the results start to deteriorate, resulting in a ratio which is far from 2.

Experiment 4.

The relative accuracy ε_r in measuring the anharmonic susceptibility,
using the non-weighted (pure) linear interpolation,
for $\lambda = 1$

N	ε_r
8	0.2959
32	0.2102
128	0.1485
512	0.1270
2048	0.1254

As expected, the non-weighted linear interpolation results are even worse than the weighted linear interpolation results in Ex. 3.

Experiment 5.

the relative accuracy ε_r in measuring the anharmonic susceptibility, with cycle index $\gamma = 2$ using the nonlinear minimization interpolation and the approximate Hamiltonian approach for $\lambda = 1$

N	χ_h	$sigma_h$	ε_r
4	0.024053	0.03077	
8	0.02535	0.03422	0.555
16	0.02604	0.03602	0.444
32	0.02655	0.03715	0.323
64	0.02695	0.03790	0.251
128	0.02727	0.03848	0.166
256	0.02751	0.03890	0.123
512	0.02769	0.03916	0.0894
1024	0.027829	0.03934	0.0661
2048	0.027925	0.039486	0.0456
4096	0.027993	0.039585	

This experiment demonstrates optimal results for a classical anharmonic crystal model Hamiltonian, since the ratio between succesive entries of ε_r in the table tends to $\sqrt{2}$ as N becomes larger. This means that application of a $W-$cycle with the nonlinear minimization interpolation and with the 4^{th} order polynomial approximation for the coarse grid Hamiltonians yield optimal results. To be more accurate, since the application of a $W-$cycle for a grid size N requires $O(N \log_2 N)$ operations, it turns out that the results are optimal upto a logarithmic factor. In order to achieve full optimality, one has to construct an appropriate multigrid cycle such that the geometric mean of the cycle indices is approximately 2.

Observe that although the relative accuracy ε_r is measured with respect to χ_h and not with respect to the desired thermodynamic limit χ_0, it presents truly the accuracy in measuring the continuum susceptibility, since the relative discretization error $\frac{|\langle\chi_h\rangle - \langle\chi_0\rangle|}{\sigma}$, is smaller than the relative accuracy ε_r.

9. Two-level performance tests

The two-level analysis allows us to test a proposed multigrid method, *without employing full multigrid*, regarding the question of the volume factor elimination. In particular, it gives us an indication about the suitability of a proposed coarsening scheme: only techniques with satisfactory two-level performance may be considered as prototypes for multigrid algorithms. This is especially useful for complicated Hamiltonians, whose form on coarse grids are not straightforwardly calculable. In the present work, we have used two-level tests in order to examine the suitability of the weighted linear interpolation and the nonlinear minimization interpolation for constructing optimal multigrid algorithms.

An exact desired fine-grid observable average, on a grid with linear size N and meshsize $h = L/N$, is denoted by A^h, and σ^h denotes the standard deviation from A^h exhibited by a single fine-grid configuration. The thermodynamic limit A is the limit of A^h as $h \to 0$. We assume that $|A - A^h|$ is smaller than $|A - A^{2h}|$. Let us examine the size of the deviation from A^h introduced by each coarsening from the relevant fine-level (meshsize h) to the next coarser level (meshsize $2h$). More precisely, we define

$$\sigma_c^h = \sqrt{\langle (A_c^h(\tilde{u}) - A^h)^2 \rangle_{\tilde{u}}}$$

where $A_c^h(\tilde{u})$ stands for the average of many (practically infinite) measurements of the desired observable on the coarse grid (e.g., after each Monte Carlo sweep on the coarse grid), starting the coarsening at fine-grid equilibrium configuration $\tilde{u}$. Averaging the deviation over an ensemble of many fine-grid equilibrium configurations yields the "*average coarsening deviation*" (ACD) for coarsening from level h to level $2h$, denoted here by σ_c^h.

The number of independent configurations that can be *usefully* averaged over after coarsening from level h to level $2h$ is

$$n_h = O\left(\left(\frac{\sigma^h}{\sigma_c^h}\right)^2\right). \tag{58}$$

Consequently, the multigrid cycle index γ should be bounded by

$$\gamma \le \frac{n_h}{n_{2h}} = \frac{(\sigma^h)^2 (\sigma_c^{2h})^2}{(\sigma^{2h})^2 (\sigma_c^h)^2},$$

otherwise extra and useless work is done without getting any additional accuracy, resulting of course in a *non*-optimal multigrid algorithm.

Moreover and perhaps more important, since for the calculation of observables that strongly depend on smooth components like susceptibility, optimal multigrid algorithms are possible only if $\gamma > 2^d$ (when most of the work is devoted to efficiently sampling on the coarse grids), the following condition should be satisfied

$$\frac{n_h}{n_{2h}} > 2^d. \tag{59}$$

In statistical optimal multigrid algorithms the fluctuation of a given mode should be approximately the same at all levels, hence $\sigma^h \approx \sigma^{2h} \approx \sigma$, where σ is the standard deviation from the continuous thermodynamic limit exhibited by each continuous configuration. Therefore, according to (58) and (59), the ACD's should satisfy for any meshsize h

$$\frac{\sigma_c^{2h}}{\sigma_c^h} > 2^{d/2},$$

which means that optimal multigrid algorithm, eliminating the volume factor, is feasible only if the two-level performance tests exhibits deviation which is smaller at least by a factor $2^{d/2}$ upon using twice finer grid.

In the following experiments the ACD was averaged over an ensemble of 1000 runs, except for Ex. 10 where the ACD was averaged over an ensemble of 2000 for $N = 4$ and over an ensemble of 100 for $N = 8, 16$.

Experiment 6.

$$\mathcal{H}_h(u^h) = \frac{\lambda}{h^3}\sum_{i=1}^{N}(u_i - u_{i-1})^4 + \frac{1}{h}\sum_{i=1}^{N}(u_i - u_{i-1})^2.$$

ACD in measuring the anharmonic susceptibility,
using the weighted linear interpolation,
for $\lambda = 1$

N	ACD
8	0.0054
16	0.0041
32	0.0030
64	0.0023
128	0.0016
256	0.0011
512	0.00080
1024	0.00056
2048	0.00041
4096	0.00027

Since the classical anharmonic crystal Hamiltonian has uniform couplings on the finest grid, the weighted linear interpolation from the next coarser level to the finest level is actually pure linear interpolation (all weights are equal to 1/2), which coincides in this case with the minimization interpolation. This experiment shows that it is feasible to construct an optimal multigrid algorithm for measuring the anharmonic susceptibility (the ratio is approximately $\sqrt{2}$), while using weighted linear interpolation between the two finest levels. Therefore, the conclusion at this stage is that the deterioration in the numerical results (Ex. 3 in Sec. 8) is a result of uneffective interpolation on the next coarser levels. This point is explained extensively in Sec. 4.1.

Experiment 7.

$$H(u) = \sum_{i=1}^{N} f_i(u_i - u_{i-1})^4 + \sum_{i=1}^{N} g_i(u_i - u_{i-1})^2$$

relative ACD in measuring the harmonic susceptibility,
using the weighted linear interpolation
for $f_i \equiv 1$, $g_{2i+1} = 1$ and $g_{2i} = 10$

N	relative ACD
4	0.0857
8	0.0348
16	0.0207
32	0.0132
64	0.00965
128	0.00698
256	0.00505
512	0.00405
1024	0.00311

Here, unlike the previous experiment, the weighted linear interpolation is not fully effective, as anticipated by the discussion in Sec. 4.1.

Experiment 8.

relative ACD in measuring the harmonic susceptibility,
using the weighted linear interpolation
for $f_i \equiv 1$, $g_{4i+1} = g_{4i+2} = 1$ and $g_{4i+3} = g_{4i} = 10$

N	relative ACD
8	0.101
16	0.0998
32	0.0682
64	0.0447
128	0.0323
256	0.0225
512	0.0159
1024	0.0114

This experiment shows promising results for a case that the weighted linear interpolation is actually the pure linear interpolation, which coincides in this case with the minimization interpolation.

Experiment 9.

relative ACD in measuring the harmonic susceptibility,
using the weighted linear interpolation
for $f_i \equiv 1$ and $g_i \equiv 1$

N	relative ACD
4	0.1725
8	0.119
16	0.0800
32	0.0561
64	0.0398
128	0.0270
256	0.0195
512	0.0133
1024	0.00951

Again, promising results are obtained for a case that the weighted linear interpolation coincides with the pure linear interpolation (which is the minimization interpolation).

Experiment 10.

Two-level unigrid tests:
ACD in measuring the harmonic susceptibility,
using the nonlinear minimization interpolation operator
$f_i \equiv 1$, $g_{2i+1} = 1$ and $g_{2i} = g = 10$

$N = 4$ and $\sigma_h = 0.0596$

number of passes on the coarse grid	ACD
10	0.0185
100	0.0078
1000	0.0064
2000	0.0063
4000	0.0059
8000	0.0059

$N = 8$ and $\sigma_h = 0.128$

number of passes on the coarse grid	ACD
$4 \cdot 10^3$	0.0031
$16 \cdot 10^3$	0.0016
$32 \cdot 10^3$	0.0011
$64 \cdot 10^3$	0.00080
$128 \cdot 10^3$	0.00058
$256 \cdot 10^3$	0.00038
$512 \cdot 10^3$	0.00026
$1024 \cdot 10^3$	0.00020
$2048 \cdot 10^3$	0.00012
$4096 \cdot 10^3$	0.000083
$8192 \cdot 10^3$	0.000078
$16384 \cdot 10^3$	0.000039
$32768 \cdot 10^3$	0.000034

$N = 16$ and $\sigma_h = 0.264$

number of passes on the coarse grid	ACD
$32 \cdot 10^3$	0.0034
$64 \cdot 10^3$	0.0019
$128 \cdot 10^3$	0.0016
$256 \cdot 10^3$	0.0012
$512 \cdot 10^3$	0.00080
$1024 \cdot 10^3$	0.00058
$2048 \cdot 10^3$	0.00038
$4096 \cdot 10^3$	0.00024

These promising results of the unigrid two-level tests for a non-trivial case of the anharmonic crystal model (actually, it simulates a typical Hamiltonian on level $k = 1$), encouraged us to develope a multigrid algorithm using the nonlinear minimization interpolation (see Sec. 7) that yielded optimal results (see Ex. 5 in Sec. 8).

In this experiment, better results are obtained than in the uniform coupling case in Ex. 9, although in both cases the minimization interpolation is used. The reason behind this peculiarity is that the application of this particular experiment with the minimization interpolation is very similar to the decimation process (which usually yields as good results as one wants). Indeed, calculation of the asymptotic coarse-grid Hamiltonian, i.e., the limit of the coarse-grid Hamiltonian as $g \to \infty$, in both processes yields the same energy action

$$\sum_{i=1}^{N/2}(u_{2i} - u_{2i-2})^4 + \sum_{i=1}^{N/2}(u_{2i} - u_{2i-2})^2.$$

In contrary, the asymptotic coarse-grid Hamiltonian for this particular experiment with the weighted linear interpolation yields, as expected, different energy action $\sum_{i=1}^{N/2}(u_{2i} - u_{2i-2})^4 + 5\sum_{i=1}^{N/2}(u_{2i} - u_{2i-2})^2$.

10. Summary

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

Multigrid algorithms can reduce and even eliminate not only the *critical slowing down* factor N^z but also the *volume factor* N^d, even in the case of the one dimensional anharmonic crystal model, where strong coupling between different scales does exist.

For the optimal calculation of the susceptibility in the one dimensional anharmonic crystal model, it is essential to use, in addition to suitable cycling, nonlinear interpolation operators which can be determined from the local Hamiltonians.

The optimal efficiency is obtained *independently* of the nonlinearity strength. The critical slowing down *and* the volume factor are completely eliminated, and the total required computational work is just $O(\varepsilon_r^{-2})$.

Acknowledgements

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 and No. 696/97 from the Israeli Academy of Science and Humanities and by the Carl F. Gauss Minerva Center for Scientific Computation.

References

1. 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 supercomputing*, S.F. McCormick, ed. (Marcel Dekker, 1988), pp. 35–62.

2. A. Brandt: Multigrid Methods in Lattice Field Computations. *Nucl. Phys.* (Proc. Suppl.) **B26**:137-180 (1992).

3. A. Brandt and M. Galun: Optimal multigrid algorithms for the massive Gaussian model and path integrals, *J. Stat. Phys.* **82**:1503–1518 (1996).

4. A. Brandt and M. Galun: Optimal multigrid algorithms for variable-coupling isotropic Gaussian models, *J. Stat. Phys.* **88**:637–664 (1997).

5. A. Brandt, M. Galun and D. Ron: Optimal multigrid algorithms for calculating thermodynamic limits, *J. Stat. Phys.* **74**:313–348 (1994).

6. 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.

7. R.C. Brower and P. Tamayo: Embedded Dynamics for ϕ^4 Theory. *Phys. Rev. Let.* **62**:1087–1090 (1989).

8. R. G. Edwards, J. Goodman and A.D. Sokal: Multi-grid Monte Carlo II. Two-dimensional XY model. *Nucl. Phys.* **354**:B289 (1991).

9. R. G. Edwards, S. J. Ferreira, J. Goodman and A.D. Sokal: Multi-grid Monte Carlo III. Two-dimensional $O(4)$-symmetric non-linear σ-model. *Nucl. Phys.* **B380**:621-664 (1992).

10. R. Esch: Functional Approximation, in *Handbook of Applied Mathematics*, C. E. Pearson, ed. (Van Nostrand Reinhold, 1983), pp. 928–987.

11. M. Galun: Optimal Multigrid Algorithms for Model Problems in Statistical Mechanics. *M.Sc. Thesis,* Weizmann Institute of Science (1992).

12. J. Goodman and A. D. Sokal: Multigrid Monte Carlo method. Conceptual foundations. *Phys. Rev.* **D40**:2034–2071 (1989).

13. J. Goodman and A.D. Sokal: Multigrid Monte Carlo methods for lattice field theories. *Phys. Rev. Lett.* **56**:1015–1018 (1986).

14. M. Hasenbusch, S. Meyer and G. Mack: Noncritical multigrid Monte Carlo: $O(3)$ non-linear σ model. *Nucl. Phy.* (Proc. Suppl.) **B20**:110-113 (1991).

15. D. Kandel, E. Domany and A. Brandt: Simulations without critical slowing down—Ising and 3-State Potts models. *Phys. Rev.* **B40**:330 (1989).

16. D. Kandel, E. Domany, D. Ron, A. Brandt and E. Loh, Jr.: Simulations without critical slowing down. *Phys. Rev. Lett.* **60**:1591 (1988).

17. D. E. Knuth: The art of computer programmings, 2nd ed. (Addison-Wesley, Reading, MA, 1981), Vol. 2, Sec. 3.4.1, pp 120 and 121; L. Devroye, Non-uniform random variate generation (Springer, New York, 1986), Sec. II.3.

18. 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).

19. G. Mack and A. Pordt: Convergent perturbation expansions for euclidean quantum field theory. *Comm. Math. Phys.* **97**:267 (1985); G. Mack, in *Non-perturbative Quantum Field Theory,* G. t'Hooft et al. (Eds.), Plenum Press, NY: 309 (1988).

20. G. Mana, A. Pelissetto and A.D. Sokal: Multigrid Monte Carlo simulation via XY Embedding. II. Two-dimensional $SU(3)$ principal chiral model. *Phys. Rev.* **D55**:3674–3741 (1997).

21. T. Mendez, A. Pelissetto and A.D. Sokal: Multi-grid Monte Carlo via XY Embedding. I. General theory and two-dimensional $O(N)$-symmetric non-linear σ-models. *Nucl. Phys.* **B477**:203–270 (1996).

22. T. Mendez and A.D. Sokal: Multi-Grid Monte Carlo. IV. One-dimensional $O(4)$-symmetric nonlinear σ-model. *Phys. Rev.* **D53**:3438–3444 (1996).

23. N. Metropolis, A.W Rosenbluth, M.N Rosenbluth, A.H Teller and E. Teller: Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**:1087 (1953).

24. R.H. Swendsen and J.S. Wang: Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58**:86–88 (1987).

25. U. Wolff: Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.* **62**:361–364 (1989).