



## Stochastic Simulation Techniques for Partition Function Approximation of Gibbs Random Field Images

*Gerasimos Potamianos and John Goutsias*

Department of Electrical and Computer Engineering  
Image Analysis and Communications Laboratory  
The Johns Hopkins University  
Baltimore, MD 21218

*Abstract* - A Monte-Carlo simulation technique for the calculation of the partition function of a general Gibbs random field is presented. We show that the partition function of a general Gibbs random field is equivalent to an expectation. This observation allows us to develop an importance sampling approach for estimating this expectation by using Monte-Carlo simulations. Two different methods are proposed for this task. We show that the resulting estimators are unbiased and consistent. Computations are performed iteratively, by using a simple, Metropolis-like, Monte-Carlo algorithm with remarkable success, as it is demonstrated by our simulations. Our work concentrates on binary, second-order Gibbs random fields defined on a rectangular lattice. However, the proposed methods can be easily extended to more general Gibbs random fields and, therefore, can become quite useful in many scientific areas, such as biology, statistical mechanics and image processing. Furthermore, a potential contribution of our technique to optimally estimating the parameters of a general Gibbs random field from a given realization via a maximum-likelihood approach is anticipated.

## I. INTRODUCTION

*Markov*, or, equivalently, *Gibbs random fields* (GRF's) belong to a well known and popular class of parametric random field models [1], [2]. They are extensively used for modeling spatial interaction phenomena, including the phenomenon of *phase transition* which occurs when some macroscopic properties of a physical system change discontinuously over a small perturbation of its parameters. Some "classic" application areas of GRF's include statistical mechanics [3-5], ecology [6], sociology [7] and crystallography [8], [9]. Recently, GRF models have found wide applicability in the general areas of image analysis and computer vision, where they have become an attractive tool for providing statistical models for images [10-13]. Since they capture various image characteristics in terms of a few parameters, they have been successfully applied in a variety of image processing problems, such as smoothing and segmentation [14-18], restoration [19], [20], reconstruction [21] and coding [22], as well as in a variety of computer vision tasks [23], [24]. However, many theoretical and computational difficulties prohibit the application of these models to a wider class of problems, one of the major difficulties being the computation of the *partition function*. No exact solutions are known for this problem, except for very simple cases [5], which are not adequate for most applications of interest. As an alternative to exact solutions, approximation methods are used to get a good estimate of this quantity. Some of the better known approximation techniques could be grouped into the following categories [5]: (a) cell, or cluster, approximations [25], [26], (b) series expansions in powers of an appropriate variable [27], [28], (c) renormalization group techniques [29], [30]; and, (d) statistical simulation and other techniques [5], [31]. Most of these methods are either unreliable, especially around the *critical point* where a phase transition occurs, or require considerable faith in the involved

assumptions [5].

The lack of a closed form solution for the partition function imposes many restrictions. For example, the statistical inference of GRF's, via a maximum-likelihood approach, is an open problem, since solution of the likelihood equations is impossible without knowing the exact value of the partition function and, possibly, some of its derivatives. As a result of this, alternative methods have been developed with moderate success [6], [11], [17], [20], [32-36]. It is, therefore, clear that the problem of calculating the partition function of a general GRF, in a computationally amenable way, is of great interest. The study of this problem is the purpose of the present paper.

The approximation technique proposed here is based on a stochastic simulation approach and it makes use of a *mutually compatible Gibbs random field* (or, equivalently, *Markov mesh random field*) and its relation to a general GRF [37], [38]. Section II is devoted to establishing the required background and notation. In Section III we discuss the problem of computing the partition function of a general GRF via Monte-Carlo simulations, and we propose two methods to achieve this, which correspond to different approximations of a general GRF by a MC-GRF random field. In Section IV we consider the algorithmic implementation of these methods and discuss various important properties of the obtained estimators. Estimation of the derivatives of the partition function, and other related quantities, is very important and is also necessary in order to carry out our second method. This is discussed in Section V, whereas, in Section VI various supporting simulation experiments are presented. Finally, in Section VII, we review our results and draw our conclusions.

## II. GIBBS RANDOM FIELDS

Assume that we have a collection of  $M \times N$  sites which form a two-dimensional rectangular lattice  $\Lambda_{MN} = \{(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$ . A discrete random variable  $H(i, j)$  is assigned at each site of the lattice, taking values from a discrete ensemble  $E_H = \{\phi_1, \phi_2, \dots, \phi_R\}$ , which contains  $R \geq 2$  distinct values. The resulting random field  $[H] = \{H(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$  can take any one of the  $R^{MN}$  possible realizations  $[h] = \{h_{ij} : 1 \leq i \leq M, 1 \leq j \leq N\}$  in the Cartesian product  $E_H^{MN}$  with joint probability distribution<sup>1</sup>  $Pr[\mathbf{H}=\mathbf{h}]$ . At each lattice site  $s_n, n = 1, 2, \dots, MN$ , a set  $N_{s_n}$  of *neighbor sites* is assigned such that: (a)  $s_n \notin N_{s_n}$ ; and, (b) if  $s_n \in N_{s_m}$ , then  $s_m \in N_{s_n}$ . We shall restrict  $[H]$  to be in the class of *Markov random fields* (MRF's). In this case the following Markovian property is satisfied:

$$\begin{aligned} Pr[H(s_n) = h(s_n) \mid H(s) = h(s), s \in \Lambda_{MN} - \{s_n\}] \\ = Pr[H(s_n) = h(s_n) \mid H(s) = h(s), s \in N_{s_n}] > 0. \end{aligned}$$

According to the Hammersley-Clifford Theorem [6],  $[H]$  is equivalent to a GRF whose joint probability distribution is given by the Gibbs measure

$$Pr[\mathbf{H} = \mathbf{h}] = \frac{1}{Z} \exp\left\{-\frac{1}{T} U(\mathbf{h})\right\}. \quad (1)$$

In eq. (1),  $Z$  is a normalizing constant known as the *partition function*,  $T$  is a positive parameter, known as the *temperature*, which controls the degree of peaking in the Gibbs measure, whereas,  $U(\cdot)$  is the *energy function* which depends on a specific realization  $\mathbf{h}$  of the GRF and the *potentials* associated with its *cliques* [6].

<sup>1</sup> In the following, boldface characters denote vectors which correspond to a lexicographic ordering of the random field.

Without loss of generality, we consider *second-order* GRF's [6]. These random fields is a "rich" class of models capable of describing a variety of spatial interaction phenomena in a satisfactory way. In this case, the neighborhood  $N_{ij}^{(2)}$ , at site  $(i, j) \in \Lambda_{MN}$ , will be given by

$$N_{ij}^{(2)} = \{(i-1, j-1), (i-1, j), (i-1, j+1), (i, j-1), (i, j+1), (i+1, j-1), (i+1, j), (i+1, j+1)\},$$

and the Gibbs measure (1) by

$$Pr[\mathbf{H} = \mathbf{h}] = \frac{1}{Z_{MN}} \prod_{i=1}^M \prod_{j=1}^N \sigma_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}), \quad (2a)$$

where

$$Z_{MN} = \sum_{\text{states } \mathbf{h}} \prod_{i=1}^M \prod_{j=1}^N \sigma_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}). \quad (2b)$$

In eq. (2),  $\sigma_{ij}(x, y, z, \omega)$  is the *local transfer function* (LTF) of the GRF  $[H]$ , and the summation is carried over all  $R^{MN}$  states [37]. The LTF has to be modified at the boundary sites of  $\Lambda_{MN}$  depending on the type of boundary conditions assumed (free, or toroidal). In this paper, we shall assume free boundary conditions<sup>2</sup>; i.e.,  $H(i, j) = \phi_1$  in  $\{(i, j), i \leq 0 \text{ or } j \leq 0\}$ . The case of a *first-order* GRF with neighborhood

$$N_{ij}^{(1)} = \{(i-1, j), (i, j-1), (i, j+1), (i+1, j)\},$$

is a special case of eq. (2), with  $\sigma_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})$  not depending on  $h_{i-1, j-1}$ .

A special case of a general GRF is a *mutually compatible Gibbs random field* (MC-GRF), or, equivalently, a *Markov mesh random field* [37]. The probability structure of a subset  $[H]_A$  of such a random field  $[H]$ , restricted on a finite sublattice  $A$  of  $\Lambda_{MN}$ , is independent of the size, or shape, of  $A$ . The LTF  $\tau_{ij}(x, y, z, \omega)$  of these GRF's is restricted to satisfy the following relationship:

<sup>2</sup> This type of boundary conditions is most natural in many practical situations [19].

$$\sum_{u \in E_H} \tau_{ij}(u, y, z, \omega) = k_{ij},$$

for every triplet  $(y, z, \omega) \in E_H^3$ , where  $k_{ij}$  is a constant. It is easy to show that the computation of the partition function  $Z_{MN}$  in this case is trivial. Indeed,

$$\begin{aligned} Z_{MN} &= \sum_{\text{states } \mathbf{h}} \prod_{i=1}^M \prod_{j=1}^N \tau_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}) \\ &= \prod_{i=1}^M \prod_{j=1}^N \sum_{u \in E_H} \tau_{ij}(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}) = \prod_{i=1}^M \prod_{j=1}^N k_{ij}. \end{aligned}$$

In this case, the partition function is expressed as the product of regular and local partition functions (the  $k_{ij}$ 's); therefore, no phase transition is associated with these models [37].

Generating a realization of a MC-GRF is also an easy task, because this can be done lexicographically using point by point simulation, as the following relation implies:

$$\begin{aligned} Pr[\mathbf{H} = \mathbf{h}] &= \prod_{i=1}^M \prod_{j=1}^N \frac{\tau_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\sum_{u \in E_H} \tau_{ij}(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \\ &= \prod_{i=1}^M \prod_{j=1}^N Pr[h_{ij} | h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}], \end{aligned}$$

where

$$Pr[h_{ij} | h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}] = \frac{\tau_{ij}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\sum_{u \in E_H} \tau_{ij}(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}.$$

Therefore, knowing the values of random variables  $H(i-1, j)$ ,  $H(i-1, j-1)$  and  $H(i, j-1)$ , we can generate the value of  $H(i, j)$  by sampling the conditional probability  $Pr[h_{ij} | h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}]$  over all possible values in  $E_H$ .

Since a MC-GRF is characterized by many attractive properties, it has been pointed out that it might be a good idea to try to approximate a general GRF by a MC-GRF [38]. Two approaches have been adopted here for the

solution of this approximation problem. The first one is the simplest one, however the second one achieves optimality with respect to an "entropy" distance between the probability distribution of a general GRF and the probability distribution of a MC-GRF. Both will be used to develop an *importance sampling* procedure for the Monte-Carlo calculation of the partition function  $Z_{MN}$ , given by eq. (2b). This will be discussed in the following section.

### III. MONTE-CARLO CALCULATION OF THE PARTITION FUNCTION

As we mentioned before, the main purpose of our work is to calculate the partition function  $Z_{MN}$  of a general GRF via *Monte-Carlo simulations*. From eq. (2b) observe that

$$Z_{MN} = \sum_{\text{states } \mathbf{h}} A_{MN}(\mathbf{h}), \quad (3a)$$

where

$$A_{MN}(\mathbf{h}) = \prod_{i=1}^M \prod_{j=1}^N \sigma_{ij}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1}). \quad (3b)$$

Since the previous summation is carried over  $R^{MN}$  states, it is not possible to compute  $Z_{MN}$  directly, even for moderate size lattices. Assume that  $P_{MN}(\mathbf{h})$  is a joint probability distribution defined on the space of all  $R^{MN}$  realizations  $\mathbf{h}$ , such that  $P_{MN}(\mathbf{h}) > 0$ , for every state  $\mathbf{h} \in E_H^{MN}$ . In this case

$$Z_{MN} = \sum_{\text{states } \mathbf{h}} \left[ \frac{A_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right] P_{MN}(\mathbf{h}) = E_P \left[ \frac{A_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right]. \quad (4)$$

From eq. (4) we derive the following estimate  $\hat{Z}_{MN,P}$  for the partition function  $Z_{MN}$ :

$$\hat{Z}_{MN,P} = \lim_{K \rightarrow \infty} \hat{Z}_{MN,P}(K) = \lim_{K \rightarrow \infty} \left[ \frac{1}{K} \sum_{k=1}^K \frac{A_{MN}(\mathbf{h}_k)}{P_{MN}(\mathbf{h}_k)} \right], \quad (5a)$$

where

$$\hat{Z}_{MN,P}(K) = \frac{1}{K} \sum_{k=1}^K \frac{A_{MN}(\mathbf{h}_k)}{P_{MN}(\mathbf{h}_k)}, \quad (5b)$$

provided that

$$\frac{A_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} < +\infty, \quad (5c)$$

for every  $\mathbf{h} \in E_H^{MN}$ . In eq. (5),  $\mathbf{h}_k$ ,  $k = 1, 2, \dots, K$ , are i.i.d. realizations drawn with probability  $P_{MN}(\mathbf{h}_k)$ , and, therefore,  $\hat{Z}_{MN,P}$  is a Monte-Carlo estimate of the partition function [39]. Inequality (5c) is satisfied in the case of finite lattices. When  $MN \rightarrow +\infty$ , this inequality may be violated and our approach has to be modified. This modification will be discussed shortly.

Let us now focus on the problem of choosing the appropriate joint probability distribution  $P_{MN}(\mathbf{h})$ . Since  $\hat{Z}_{MN,P}(K)$ , and, therefore,  $\hat{Z}_{MN,P}$ , are unbiased estimators of  $Z_{MN}$  [39], we shall concentrate on finding joint probability distributions  $P_{MN}(\mathbf{h})$  which result in a small error variance  $E_P[(\hat{Z}_{MN,P}(K) - Z_{MN})^2]$ , for every  $K$ . Ultimately, we would like to minimize the quantity

$$E_P[\hat{Z}_{MN,P}^2(K)] = \frac{1}{K} \sum_{\text{states } \mathbf{h}} \frac{A_{MN}^2(\mathbf{h})}{P_{MN}(\mathbf{h})} + \frac{K-1}{K} Z_{MN}^2,$$

with respect to  $P_{MN}(\mathbf{h})$ , provided that we can draw samples  $\mathbf{h}$  from the probability distribution  $P_{MN}(\mathbf{h})$  in a computationally efficient way. Equivalently, we would like to minimize (with respect to  $P_{MN}(\mathbf{h})$ ) the *Ali-Silvey distance* [40]

$$\begin{aligned} d_{IS}(\pi_{MN}, P_{MN}) &= \sum_{\text{states } \mathbf{h}} \left[ \frac{\pi_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right]^2 P_{MN}(\mathbf{h}) \\ &= \frac{K}{Z_{MN}^2} E_P[\hat{Z}_{MN,P}^2(K)] - (K-1), \end{aligned} \quad (6)$$

between the probability measures  $P_{MN}(\mathbf{h})$  and  $\pi_{MN}(\mathbf{h})$ . Notice that  $P_{MN}(\mathbf{h})$  should be known explicitly, since at each step of the Monte-Carlo algorithm the quantity  $A_{MN}(\mathbf{h})/P_{MN}(\mathbf{h})$  must be calculated.

It is straightforward to show that the direct minimization of eq. (6), with respect to  $P_{MN}(\mathbf{h})$ , results in  $P_{MN}(\mathbf{h}) = \pi_{MN}(\mathbf{h}) = A_{MN}(\mathbf{h}) / Z_{MN}$ , which is the Gibbs measure. Although this choice results in zero error variance, it is useless, since the computation of  $A_{MN}(\mathbf{h}) / P_{MN}(\mathbf{h})$  requires  $Z_{MN}$  to be known explicitly. We can resolve this problem by considering probability distributions  $P_{MN}(\mathbf{h})$  which correspond to a MC-GRF, therefore, satisfying the following relation:

$$P_{MN}(\mathbf{h}) = \prod_{i=1}^M \prod_{j=1}^N \tau_{ij}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1}) > 0, \quad (7a)$$

where

$$\sum_{u \in E_H} \tau_{ij}(u, y, z, \omega) = 1, \text{ for every } (y, z, \omega) \in E_H^3. \quad (7b)$$

We shall denote this class of distributions by  $\mathbf{D}$ ; i.e.,

$$P_{MN}(\mathbf{h}) \in \mathbf{D}, \quad \text{if and only if eq. (7) holds.}$$

This is a convenient choice, since drawing samples from  $P_{MN}(\mathbf{h})$  (i.e., generating a realization of a MC-GRF) is an easy task, as it was mentioned in the previous section. Computing the ratio  $A_{MN}(\mathbf{h}) / P_{MN}(\mathbf{h})$  is also straightforward. Indeed, if

$$q_{ij,P}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1}) = \frac{\sigma_{ij}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1})}{\tau_{ij}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1})}, \quad (8a)$$

and

$$Q_{MN,P}(\mathbf{h}) = \prod_{i=1}^M \prod_{j=1}^N q_{ij,P}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1}), \quad (8b)$$

then (see eqs. (3), (4), (7) and (8))

$$Z_{MN} = \sum_{\text{states } \mathbf{h}} Q_{MN,P}(\mathbf{h}) P_{MN}(\mathbf{h}), \quad (9)$$

and

$$\hat{Z}_{MN,P} = \lim_{K \rightarrow +\infty} \hat{Z}_{MN,P}(K) = \lim_{K \rightarrow +\infty} \left[ \frac{1}{K} \sum_{k=1}^K Q_{MN,P}(\mathbf{h}_k) \right], \quad (10a)$$

where

$$\hat{Z}_{MN,P}(K) = \frac{1}{K} \sum_{k=1}^K Q_{MN,P}(\mathbf{h}_k), \quad (10b)$$

provided that

$$Q_{MN,P}(\mathbf{h}) < +\infty, \quad (11)$$

for every  $\mathbf{h} \in E_H^{MN}$ . The error variance in this case will be given by (see eq. (10))

$$\text{Var}_P[\hat{Z}_{MN,P}(K)] = \frac{1}{K} \left[ \left[ \sum_{\text{states } \mathbf{h}} Q_{MN,P}^2(\mathbf{h}) P_{MN}(\mathbf{h}) \right] - Z_{MN}^2 \right]. \quad (12)$$

A simple choice for the LTF of a MC-GRF could be  $\tau_{ij}^{iid}(x, y, z, \omega) = 1/R$ , for every  $(x, y, z, \omega) \in E_H^4$  and every  $(i, j) \in \Lambda_{MN}$ . In this case  $P_{MN}^{iid}(\mathbf{h}) = 1/R^{MN}$  (i.e., the joint probability distribution of an i.i.d. random field) and the error variance will be given by

$$\text{Var}_{P^{iid}}[\hat{Z}_{MN,P^{iid}}(K)] = \frac{1}{K} \left[ R^{MN} \left[ \sum_{\text{states } \mathbf{h}} A_{MN}^2(\mathbf{h}) \right] - Z_{MN}^2 \right]. \quad (13)$$

In most practical situations one does not expect the i.i.d. choice to give a good estimate of  $Z_{MN}$  in a reasonable time. The reason for this is that, in most cases of interest (e.g., in low temperatures), only a small fraction of realizations  $\mathbf{h}_k$  contribute substantially to the partition function sum, while the contribution of most other realizations is negligible. Considering samples  $\mathbf{h}_k$  with equal probability will greatly underestimate  $Z_{MN}$ , since most of the time  $\mathbf{h}_k$  will be a sample that does not contribute much to the computation of  $Z_{MN}$ . To overcome this difficulty, we will need many iterations, which could result in a  $K$  being comparable to  $R^{MN}$ . Nevertheless, this sampling scheme becomes attractive in the case of GRF's which are "close" to i.i.d. random fields (i.e., in high temperatures).

If we now choose the joint probability distribution of a MC-GRF which is "as close as possible" to the Gibbs measure, then this choice will favor the most

probable realizations of the GRF over the less probable ones. This will result in the error variance given by eq. (12) which will be much less than the error variance given by eq. (13). It is now natural to seek the optimal joint probability distribution  $P_{MN}^{opt}(\mathbf{h})$  of a MC-GRF, by solving the following constrained minimization problem:

$$P_{MN}^{opt}(\mathbf{h}) = \arg \left\{ \min_{P_{MN} \in \mathbf{D}} d_{IS}(\pi_{MN}, P_{MN}) \right\}. \quad (14)$$

However, the solution of eq. (14) is not feasible in general. Therefore, we shall derive two *suboptimal* solutions to this problem. The first one is motivated by the following theorem:

*Theorem 1:* For every joint probability distribution  $P_{MN}^{(1)}(\mathbf{h}) \in \mathbf{D}$ , there will be a joint probability distribution  $P_{MN}^{(2)}(\mathbf{h}) \in \mathbf{D}$ , with

$$\tau_{ij}^{(2)}(x, y, z, \omega) = \tau_{ij}^{(1)}(x, y, z, \omega), \quad (15a)$$

for every  $(i, j) \in \Lambda_{MN} - \{(M, N)\}$  and

$$\tau_{MN}^{(2)}(x, y, z, \omega) = \frac{\sigma_{MN}(x, y, z, \omega)}{\sum_{u \in E_H} \sigma_{MN}(u, y, z, \omega)}, \quad (15b)$$

for every  $(x, y, z, \omega) \in E_H^4$ , such that  $\text{Var}_{P^{(2)}}[\hat{Z}_{MN, P^{(2)}}(K)] \leq \text{Var}_{P^{(1)}}[\hat{Z}_{MN, P^{(1)}}(K)]$ .

*Proof:* For every  $(y, z, \omega) \in E_H^3$ , let  $\tau_{MN}^{(2)}(x, y, z, \omega)$  be the LTF which minimizes the following constrained minimization problem:

$$\text{minimize } \sum_{u \in E_H} \frac{\sigma_{MN}^2(u, y, z, \omega)}{\tau_{MN}(u, y, z, \omega)}, \quad (16a)$$

such that

$$\sum_{u \in E_H} \tau_{MN}(u, y, z, \omega) = 1, \quad (16b)$$

and

$$\tau_{MN}(x, y, z, \omega) > 0, \text{ for every } x \in E_H. \quad (16c)$$

By using a Lagrange multiplier  $\lambda_{yzo}$  and by differentiating eq. (16a) with respect to  $\tau_{MN}(x, y, z, \omega)$ , we obtain

$$\frac{\partial}{\partial \tau_{MN}(x, y, z, \omega)} \left[ \sum_{u \in E_H} \frac{\sigma_{MN}^2(u, y, z, \omega)}{\tau_{MN}(u, y, z, \omega)} + \lambda_{yzo} \left[ \sum_{u \in E_H} \tau_{MN}(u, y, z, \omega) - 1 \right] \right] = 0,$$

or

$$\tau_{MN}(x, y, z, \omega) = \frac{\sigma_{MN}(x, y, z, \omega)}{\sqrt{\lambda_{yzo}}} > 0. \quad (17)$$

From eq. (16b) we have that

$$\sqrt{\lambda_{yzo}} = \sum_{u \in E_H} \sigma_{MN}(u, y, z, \omega),$$

which together with eq. (17) gives eq. (15b). Notice that, the only non-zero entries of the Hessian of

$$\sum_{u \in E_H} \frac{\sigma_{MN}^2(u, y, z, \omega)}{\tau_{MN}(u, y, z, \omega)} + \lambda_{yzo} \left[ \sum_{u \in E_H} \tau_{MN}(u, y, z, \omega) - 1 \right]$$

will be the diagonal ones, which are positive. Therefore, the LTF  $\tau_{MN}^{(2)}(x, y, z, \omega)$ , given by eq. (15b), is the solution of problem (16). If  $\Lambda'_{MN} = \Lambda_{MN} - \{(M, N)\}$ , we have that (see also eqs. (7), (8), (12) and (15))

$$\begin{aligned} K \text{Var}_{P(1)} \left[ \hat{Z}_{MN, P(1)}(K) \right] + Z_{MN}^2 &= \sum_{\text{states } h \text{ on } \Lambda_{MN}} \prod_{i=1}^M \prod_{j=1}^N \frac{\sigma_{ij}^2(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{ij}^{(1)}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \\ &= \sum_{\text{states } h \text{ on } \Lambda'_{MN}} \left[ \prod_{(i, j) \in \Lambda'_{MN}} \frac{\sigma_{ij}^2(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{ij}^{(1)}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \left[ \sum_{u \in E_H} \frac{\sigma_{MN}^2(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{MN}^{(1)}(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \right] \right] \\ &\geq \sum_{\text{states } h \text{ on } \Lambda'_{MN}} \left[ \prod_{(i, j) \in \Lambda'_{MN}} \frac{\sigma_{ij}^2(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{ij}^{(2)}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \left[ \sum_{u \in E_H} \frac{\sigma_{MN}^2(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{MN}^{(2)}(u, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} \right] \right] \\ &= \sum_{\text{states } h \text{ on } \Lambda_{MN}} \prod_{i=1}^M \prod_{j=1}^N \frac{\sigma_{ij}^2(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{\tau_{ij}^{(2)}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})} = K \text{Var}_{P(2)} \left[ \hat{Z}_{MN, P(2)}(K) \right] + Z_{MN}^2. \quad (18) \end{aligned}$$

The proof is now a direct consequence of eq. (18). □

Given the LTF  $\sigma_{ij}(x, y, z, \omega)$  of a GRF, our first suboptimal choice for the joint probability distribution  $P_{MN}(\mathbf{h})$  will be a joint probability distribution  $P_{MN}^*(\mathbf{h})$ , given by eq. (7), with LTF

$$\tau_{ij}^*(x, y, z, \omega) = \frac{\sigma_{ij}(x, y, z, \omega)}{\sum_{u \in E_H} \sigma_{ij}(u, y, z, \omega)}, \quad (19)$$

for every  $(x, y, z, \omega) \in E_H^4$  and every  $(i, j) \in \Lambda_{MN}$ . In this case (see eq. (8a))

$$q_{ij, P^*}(x, y, z, \omega) = \sum_{u \in E_H} \sigma_{ij}(u, y, z, \omega). \quad (20)$$

The MC-GRF with LTF given by eq. (19) corresponds to the MC-GRF obtained by approximating a general GRF via the "Approach B" developed in [38]. This is a sensible choice, because the computation of the joint probability distribution  $P_{MN}^*(\mathbf{h})$  is feasible, and the obtained MC-GRF contains substantial information about the original GRF [38], much more than the i.i.d. random field, which contains absolutely no information about the GRF in low temperatures. One expects that the samples  $\mathbf{h}$ , drawn from this joint probability distribution, will contribute substantially to the computation of the partition function sum, resulting in a considerable variance reduction. In general, we expect that

$$\text{Var}_{P^*}[\hat{Z}_{MN, P^*}(K)] \ll \text{Var}_{P^{iid}}[\hat{Z}_{MN, P^{iid}}(K)],$$

where

$$\text{Var}_{P^*}[\hat{Z}_{MN, P^*}(K)] = \frac{1}{K} \left[ \left[ \sum_{\text{states } \mathbf{h}} \frac{A_{MN}^2(\mathbf{h})}{P_{MN}^*(\mathbf{h})} \right] - Z_{MN}^2 \right],$$

which has been verified to be the case in most of our simulations.

Probability  $P_{MN}^*(\mathbf{h})$ , although simple to compute, may result in an unacceptably high error variance. This is typically the case for GRF's in low temperatures, as our simulations demonstrate; therefore, a better choice for  $P_{MN}(\mathbf{h})$  is needed, which will result in larger error variance reduction. This is provided by a theorem, which appears in [41].

Consider the following "entropy" distance between the Gibbs distribution and a probability distribution  $P_{MN}(\mathbf{h})$ :

$$d_{IL}(\pi_{MN}, P_{MN}) = \sum_{states \mathbf{h}} \left[ \frac{\pi_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right] \ln \left[ \frac{\pi_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right] P_{MN}(\mathbf{h}) = \sum_{states \mathbf{h}} \pi_{MN}(\mathbf{h}) \ln \left[ \frac{\pi_{MN}(\mathbf{h})}{P_{MN}(\mathbf{h})} \right].$$

This is also an *Ali-Silvey* type of distance, and is minimized for  $P_{MN}(\mathbf{h}) = \pi_{MN}(\mathbf{h})$ , for all states  $\mathbf{h}$ . For the reasons stated above we would like to constrain the minimization problem and obtain a probability distribution  $P_{MN}^{**}(\mathbf{h})$  which satisfies

$$P_{MN}^{**}(\mathbf{h}) = arg \left\{ \min_{P_{MN} \in \mathbf{D}} d_{IL}(\pi_{MN}, P_{MN}) \right\}. \quad (21)$$

If we assume that the Gibbs distribution  $\pi_{MN}(\mathbf{h})$  has homogeneous<sup>3</sup> LTF's, i.e.,

$$\sigma_{ij}(x, y, z, \omega) = \sigma(x, y, z, \omega),$$

for every  $(i, j) \in \Lambda_{MN}$  and  $(x, y, z, \omega) \in E_H^4$ , then we obtain the following theorem.

*Theorem 2:* The probability measure  $P_{MN}^{**}(\mathbf{h}) \in \mathbf{D}$  with LTF

$$\tau_{ij}^{**}(x, y, z, \omega) = \frac{\left[ \frac{\partial \ln Z_{MN}}{\partial \sigma(x, y, z, \omega)} \right] \sigma(x, y, z, \omega)}{\sum_{u \in E_H} \left[ \frac{\partial \ln Z_{MN}}{\partial \sigma(u, y, z, \omega)} \right] \sigma(u, y, z, \omega)}, \quad (22)$$

for every  $(x, y, z, \omega) \in E_H^4$  and  $(i, j) \in \Lambda_{MN}$  satisfies eq. (21).

<sup>3</sup> This is a reasonable assumption in image processing applications and simplifies our derivations considerably.

*Proof:* The proof can be found in [41].

□

The MC-GRF with probability measure  $P_{MN}^{**}$  is optimal, achieving a minimum entropy distance from the Gibbs distribution  $\pi_{MN}$ . However, in general, it is *suboptimal* with respect to minimizing the error variance, or equivalently the distance  $d_{IS}(\pi_{MN}, P_{MN})$ .

Given the LTF  $\sigma(x, y, z, \omega)$  of a GRF, our second choice for the joint probability distribution  $P_{MN}(\mathbf{h})$  will be a joint probability distribution  $P_{MN}^{**}(\mathbf{h})$  given by eq. (7), with LTF given by eq. (22). In this case (see eq. (8a))

$$q_{ij, P^{**}}(x, y, z, \omega) = \frac{\sum_{u \in E_H} \left[ \frac{\partial \ln Z_{MN}}{\partial \sigma(u, y, z, \omega)} \right] \sigma(u, y, z, \omega)}{\left[ \frac{\partial \ln Z_{MN}}{\partial \sigma(x, y, z, \omega)} \right]} . \quad (23)$$

As expected, the MC-GRF with LTF given by eq. (22) approximates the original GRF better than the MC-GRF with LTF given by eq. (19), especially in low temperatures. This may result in significant variance reduction; i.e., we expect

$$\text{Var}_{P^{**}}[\hat{Z}_{MN, P^{**}}(K)] \ll \text{Var}_{P^*}[\hat{Z}_{MN, P^*}(K)] ,$$

where

$$\text{Var}_{P^{**}}[\hat{Z}_{MN, P^{**}}(K)] = \frac{1}{K} \left[ \left[ \sum_{\text{states } \mathbf{h}} \frac{A_{MN}^2(\mathbf{h})}{P_{MN}^{**}(\mathbf{h})} \right] - Z_{MN}^2 \right] ,$$

which has been verified by our simulation experiments for a variety of GRF's.

To summarize, we propose two methods for calculating the partition function  $Z_{MN}$ :

$$\text{Method 1: } Z_{MN} = E_{P^*} \left[ \frac{A_{MN}(\mathbf{h})}{P_{MN}^*(\mathbf{h})} \right] \Rightarrow \hat{Z}_{MN, P^*} = \lim_{K \rightarrow \infty} \left[ \frac{1}{K} \sum_{k=1}^K Q_{MN, P^*}(\mathbf{h}_k) \right] ,$$

where  $\mathbf{h}_k \sim P_{MN}^*(\mathbf{h}_k)$  ,

$$\text{Method 2: } Z_{MN} = E_{P^{**}} \left[ \frac{A_{MN}(\mathbf{h})}{P_{MN}^{**}(\mathbf{h})} \right] \Rightarrow \hat{Z}_{MN, P^{**}} = \lim_{K \rightarrow +\infty} \left[ \frac{1}{K} \sum_{k=1}^K Q_{MN, P^{**}}(\mathbf{h}_k) \right],$$

where  $\mathbf{h}_k \sim P_{MN}^{**}(\mathbf{h}_k)$ ,

with  $P_{MN}^*(\mathbf{h})$  and  $P_{MN}^{**}(\mathbf{h})$  being the joint probability distributions of two MC-GRF's with LTF's given by eq. (19) and eq. (22) respectively, and  $Q_{MN, P^*}(\mathbf{h})$ ,  $Q_{MN, P^{**}}(\mathbf{h})$  by eqs. (8b), (20) and (8b), (23) respectively.

#### IV. COMPUTATIONAL ALGORITHMS

After deciding for the MC-GRF joint probability distribution to be used in the Monte-Carlo partition function calculation, the next step is to determine how to implement this calculation in a computationally efficient way. In the following we shall denote both  $P_{MN}^*(\mathbf{h})$  and  $P_{MN}^{**}(\mathbf{h})$  as  $P_{MN}^\gamma(\mathbf{h})$ , i.e, the superscript  $\gamma$  will stand for \* or \*\*. A simple approach is to use the following algorithm:

##### ALGORITHM I:

1. Draw  $K$  statistically independent realizations  $\mathbf{h}_k$ ,  $k = 1, 2, \dots, K$ , from probability  $P_{MN}^\gamma(\mathbf{h})$ , given by eq. (7), with  $\tau_{ij}(x, y, z, \omega) = \tau_{ij}^\gamma(x, y, z, \omega)$ , given by eq. (19) or (22). This can be done lexicographically, as it has been discussed in Section II.
2. Compute  $Q_{MN, P^\gamma}(\mathbf{h}_k)$ , for every realization  $\mathbf{h}_k$ ,  $k = 1, 2, \dots, K$ , by using eq. (8b) and one of (20), (23).
3. Compute  $\hat{Z}_{MN, P^\gamma}(K)$  by using eq. (10b).

The statistical properties of the estimator  $\hat{Z}_{MN, P^\gamma}(K)$  can be easily derived. Indeed,  $\hat{Z}_{MN, P^\gamma}(K)$  is a sum of the i.i.d. random variables  $Q_{MN, P^\gamma}(\mathbf{h}_1)$ ,  $Q_{MN, P^\gamma}(\mathbf{h}_2)$ ,  $\dots$ ,  $Q_{MN, P^\gamma}(\mathbf{h}_K)$  divided by  $K$ . Each of these random variables has finite mean

$\mu_{MN, P\gamma} = E_{P\gamma}[Q_{MN, P\gamma}(\mathbf{h}_k)] = Z_{MN}$  and finite variance  $\sigma_{MN, P\gamma}^2 = \text{Var}_{P\gamma}[Q_{MN, P\gamma}(\mathbf{h}_k)]$   
 $= \sum_{\text{states } \mathbf{h}} Q_{MN, P\gamma}^2(\mathbf{h}) P_{MN}^{\gamma}(\mathbf{h}) - Z_{MN}^2 > 0$ . From the *central limit theorem* we have that  
 $\hat{Z}_{MN, P\gamma}(K)$  is an asymptotically normal random variable with

$$\sqrt{K} \frac{\hat{Z}_{MN, P\gamma}(K) - Z_{MN}}{\sigma_{MN, P\gamma}} \rightarrow N(0, 1),$$

as  $K \rightarrow +\infty$ . From the *strong law of large numbers* we also have that

$$\text{Pr} \left[ \lim_{K \rightarrow +\infty} \hat{Z}_{MN, P\gamma}(K) = Z_{MN} \right] = 1;$$

i.e., our estimator converges to the partition function with probability one. Therefore, it converges in probability; i.e.,

$$\text{Pr} \left[ |\hat{Z}_{MN, P\gamma}(K) - Z_{MN}| \geq \varepsilon \right] \rightarrow 0, \quad (24)$$

as  $K \rightarrow +\infty$ , for every  $\varepsilon > 0$ . Hence,  $\hat{Z}_{MN, P\gamma}(K)$  is an unbiased and consistent estimator of  $Z_{MN}$ . To get a practical idea on the accuracy of such an estimator, it is worthwhile calculating the sample variance, defined by

$$\hat{\sigma}_{MN, P\gamma}^2(K) = \frac{1}{K} \sum_{k=1}^K [Q_{MN, P\gamma}(\mathbf{h}_k) - \bar{Q}_{MN, P\gamma}(K)]^2, \quad (25a)$$

where

$$\bar{Q}_{MN, P\gamma}(K) = \frac{1}{K} \sum_{k=1}^K Q_{MN, P\gamma}(\mathbf{h}_k). \quad (25b)$$

This is the usual estimator for the variance of  $\hat{Z}_{MN, P\gamma}(K)$ , and can be effectively used to decide how many iterations  $K$  are needed in order to achieve a prespecified accuracy in estimating  $Z_{MN}$ .

The computational complexity of *Algorithm I* is  $O(KMN)$ , since it requires the generation of  $KMN$  samples drawn from the discrete conditional probability distribution  $\text{Pr}[h_{ij} | h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1}]$ ; therefore, for a large lattice, we may not be able to obtain accurate estimates of  $Z_{MN}$  in a reasonable time. This

problem may be partially solved by considering realizations  $\mathbf{h}_1, \mathbf{h}_2, \dots$ , which differ only slightly from each other. In this case,  $\mathbf{h}_{k+1}$  and  $Q_{MN, P^\gamma}(\mathbf{h}_{k+1})$  may be easily calculated from  $\mathbf{h}_k$  and  $Q_{MN, P^\gamma}(\mathbf{h}_k)$  by a simple update. This can be done by generating a *Markov chain* on the  $R^{MN}$  possible realizations of  $[H]$  with prespecified transition probabilities  $P_{ij}^{(k, k+1)}$ , given by

$$P_{ij}^{(k, k+1)} = Pr [ \mathbf{H}_{k+1} = \mathbf{h}_j \mid \mathbf{H}_k = \mathbf{h}_i ],$$

from state  $\mathbf{h}_i$ , at step  $k$ , to state  $\mathbf{h}_j$ , at step  $k + 1$ . These transition probabilities form a sequence of  $R^{MN} \times R^{MN}$  transition probability matrices  $\mathbf{P}^{(k)}$ ,  $k = 1, 2, \dots$ , which should satisfy the following relationship:

$$(\underline{P}^\gamma)^t \mathbf{P}^{(k)} = (\underline{P}^\gamma)^t, \quad (26)$$

where  $\underline{P}^\gamma$  is the  $R^{MN}$ -dimensional probability vector with elements  $P_i^\gamma = P_{MN}^\gamma(\mathbf{h}_i)$  and  $t$  denotes transposition. If  $Pr[ \mathbf{H}_1 = \mathbf{h} ] = P_{MN}^\gamma(\mathbf{h})$ , and if eq. (26) is satisfied for every  $k = 1, 2, \dots$ , then every state of our Markov chain will be a sample drawn from  $P_{MN}^\gamma(\mathbf{h})$ . Therefore, we have to construct  $\mathbf{P}^{(k)}$  such that eq. (26) is satisfied. However, we have to keep in mind that, in this case, the realizations  $\mathbf{h}_1, \mathbf{h}_2, \dots$ , are no longer statistically independent, and a new analysis is required for the study of the statistical properties of  $\hat{Z}_{MN, P^\gamma}(K)$ .

A simple way to accomplish the previous ideas is to consider a sequence  $\mathbf{h}_1, \mathbf{h}_2, \dots$ , of realizations such that  $\mathbf{h}_{k+1}$  differs from  $\mathbf{h}_k$  only at one site  $s_p$ , this site being chosen randomly (among all possible  $MN$  sites in  $\Lambda_{MN}$ ), or, systematically (e.g., lexicographically). Furthermore, the random variable  $H_{k+1}(s_p)$ , at site  $s_p$ , takes a value  $h(s_p)$  in  $E_H$  drawn from a given probability distribution. When this probability distribution is given by

$$Pr [ H_{k+1}(s_p) = h(s_p) \mid H_k(s) = h(s), s \in N_{s_p} ] > 0,$$

the resulting algorithm is known as the *Gibbs sampler* [19]. Other algorithms of a similar nature are described in [41] and are known as the *Metropolis'*,

*Barker's and Hastings' algorithms.*

We have compared the performances of all these algorithms, using random, or lexicographic, site updating. As expected from the analysis in [41], the Gibbs sampler with lexicographic site updating gave the best results: it converged, on the average, faster than the other algorithms, requiring less computational time, and resulted in lower *rejection rates* (higher percentage of successful updates). This result comes to no surprise, since the Gibbs sampler is an importance sampling based procedure [39], [42], [43].

The Gibbs sampler is the sampling scheme used exclusively in our simulations and results in the following algorithm for the estimation of  $Z_{MN}$ .

**ALGORITHM II:**

1. Generate a realization  $\mathbf{h}_1$  of the MC-GRF with probability  $P_{MN}^Y(\mathbf{h})$ , given by eq. (7), with  $\tau_{ij}(x, y, z, \omega) = \tau_{ij}^Y(x, y, z, \omega)$ , given by eq. (19) or (22). This is done lexicographically, as discussed in Section II. Calculate  $Q_{MN, P^Y}(\mathbf{h}_1)$ , given by eqs. (8b) and (20) or (8b) and (23).
2. Set  $SUM_1 = QFUN_1 = Q_{MN, P^Y}(\mathbf{h}_1)$  and  $k = 1$ .
3. Set  $h = h_k(s_p)$ , with  $p = (k - 1) \text{ modulo } MN + 1$ , where  $s_1, s_2, \dots, s_{MN}$ , is a lexicographic (row by row) ordering of the sites in  $\Lambda_{MN}$ . Denote site  $s_p$  by  $(i, j)$ .
4. Draw a value  $\phi \in E_H$  from probability  $d_n / (\sum_{l=1}^R d_l)$ ,  $n = 1, 2, \dots, R$ , where

$$d_l = \tau_{ij}^Y(\phi_l, h_{i-1, j}^{(k)}, h_{i-1, j-1}^{(k)}, h_{i, j-1}^{(k)}) \times \tau_{i, j+1}^Y(h_{i, j+1}^{(k)}, h_{i-1, j+1}^{(k)}, h_{i-1, j}^{(k)}) \\ \times \tau_{i+1, j}^Y(h_{i+1, j}^{(k)}, \phi_l, h_{i, j-1}^{(k)}, h_{i+1, j-1}^{(k)}) \times \tau_{i+1, j+1}^Y(h_{i+1, j+1}^{(k)}, h_{i, j+1}^{(k)}, \phi_l, h_{i+1, j}^{(k)}),$$

for  $l = 1, 2, \dots, R$ .

5. If  $\phi = h$  set  $QFUN_{k+1} \leftarrow QFUN_k$  and go to step 6, otherwise compute the ratio

$$RATIO = Q_{MN, P^Y}(\mathbf{h}_{k+1}) / Q_{MN, P^Y}(\mathbf{h}_k),$$

by

$$\begin{aligned}
 \text{RATIO} &= \frac{q_{i,j+1,p\gamma}(h_{i,j+1}^{(k)}, h_{i-1,j+1}^{(k)}, h_{i-1,j}^{(k)}, \phi) \times q_{i+1,j,p\gamma}(h_{i+1,j}^{(k)}, \phi, h_{i,j-1}^{(k)}, h_{i+1,j-1}^{(k)})}{q_{i,j+1,p\gamma}(h_{i,j+1}^{(k)}, h_{i-1,j+1}^{(k)}, h_{i-1,j}^{(k)}, h) \times q_{i+1,j,p\gamma}(h_{i+1,j}^{(k)}, h, h_{i,j-1}^{(k)}, h_{i+1,j-1}^{(k)})} \\
 &\times \frac{q_{i+1,j+1,p\gamma}(h_{i+1,j+1}^{(k)}, h_{i,j+1}^{(k)}, \phi, h_{i+1,j}^{(k)}) \times q_{ij,p\gamma}(\phi, h_{i-1,j}^{(k)}, h_{i-1,j-1}^{(k)}, h_{i,j-1}^{(k)})}{q_{i+1,j+1,p\gamma}(h_{i+1,j+1}^{(k)}, h_{i,j+1}^{(k)}, h, h_{i+1,j}^{(k)}) \times q_{ij,p\gamma}(h, h_{i-1,j}^{(k)}, h_{i-1,j-1}^{(k)}, h_{i,j-1}^{(k)})},
 \end{aligned}$$

where  $q_{ij,p\gamma}(x,y,z,\omega)$  is given by eq. (20) or (23). Set  $QFUN_{k+1} \leftarrow QFUN_k \times \text{RATIO}$ .

6. Set  $SUM_{k+1} \leftarrow SUM_k + QFUN_{k+1}$ . If  $k+1 = K$ , set  $\hat{Z}_{MN,p\gamma}(K) = \frac{SUM_{k+1}}{k+1}$  and stop; otherwise, set  $h_{ij}^{(k+1)} = \phi$ ,  $h_{mn}^{(k+1)} = h_{mn}^{(k)}$ , for every  $(m,n) \neq (i,j)$ ,  $k \leftarrow k+1$ , and go to step 3.

We would like now to emphasize that, in the case of *Algorithm II*, consecutive realizations of the Markov chain differ in at most one site and, therefore, they cannot "cover" rapidly the entire state space of  $[H]$ , especially in the case of large lattices. A remedy for this problem would be to consider  $L \ll MN$  successive realizations of our algorithm as one state transition of the Markov chain. In practice however, we have noticed that this modification does not improve the convergence rate of *Algorithm II*.

Once again, we are interested in the statistical properties of the estimator  $\hat{Z}_{MN,p\gamma}(K)$ , resulting from *Algorithm II*. In order to simplify our presentation, we shall consider *Algorithm II* with random (instead of lexicographic) site updating, in which case, the underlying Markov chain is characterized by stationary transition probabilities. A more general treatment (which is necessary when lexicographic site updating is considered) can be found in [44]. Both cases result in the same properties for the estimator  $\hat{Z}_{MN,p\gamma}(K)$ . As in the case of *Algorithm I*, the estimator  $\hat{Z}_{MN,p\gamma}(K)$  is the sum of identically distributed random variables  $Q_{MN,p\gamma}(\mathbf{h}_k)$ , which are characterized by a finite mean

$\mu_{MN,P\gamma} = Z_{MN}$  and finite, nonzero variance  $\sigma_{MN,P\gamma}^2$ . Obviously,  $\hat{Z}_{MN,P\gamma}(K)$  is an unbiased estimator of  $Z_{MN}$ . The computation of the variance of  $\hat{Z}_{MN,P\gamma}(K)$  requires more effort. Observe that

$$\begin{aligned} \text{Var}_{P\gamma}[\hat{Z}_{MN,P\gamma}(K)] &= E_{P\gamma} \left[ \left( \frac{1}{K} \sum_{k=1}^K Q_{MN,P\gamma}(\mathbf{h}_k) - Z_{MN} \right)^2 \right] = \frac{1}{K^2} E_{P\gamma} \left[ \left( \sum_{k=1}^K (Q_{MN,P\gamma}(\mathbf{h}_k) - Z_{MN}) \right)^2 \right] \\ &= \frac{1}{K^2} \left[ \sum_{k=1}^K E_{P\gamma}[(Q_{MN,P\gamma}(\mathbf{h}_k) - Z_{MN})^2] + 2 \sum_{m \neq n} E_{P\gamma}[(Q_{MN,P\gamma}(\mathbf{h}_m) - Z_{MN})(Q_{MN,P\gamma}(\mathbf{h}_n) - Z_{MN})] \right], \end{aligned}$$

or

$$\begin{aligned} &K \text{Var}_{P\gamma}[\hat{Z}_{MN,P\gamma}(K)] \\ &= E_{P\gamma} \left[ (Q_{MN,P\gamma}(\mathbf{h}_1) - Z_{MN})^2 \right] + 2 \sum_{l=1}^{K-1} \frac{K-l}{K} \left[ E_{P\gamma} [Q_{MN,P\gamma}(\mathbf{h}_1) Q_{MN,P\gamma}(\mathbf{h}_l)] - Z_{MN}^2 \right] \\ &= \sum_{\text{states } \mathbf{h}_m} \sum_{\text{states } \mathbf{h}_n} Q_{MN,P\gamma}(\mathbf{h}_m) Q_{MN,P\gamma}(\mathbf{h}_n) \left[ P_{MN}^{\gamma}(\mathbf{h}_m) \delta_{mn} - P_{MN}^{\gamma}(\mathbf{h}_m) P_{MN}^{\gamma}(\mathbf{h}_n) \right. \\ &\quad \left. + 2 \sum_{l=1}^{K-1} \frac{K-l}{K} P_{MN}^{\gamma}(\mathbf{h}_m) [P_{mn}^{(1,l)} - P_{MN}^{\gamma}(\mathbf{h}_n)] \right], \quad (27) \end{aligned}$$

where  $\delta_{mn}$  is the Kronecker delta, and

$$P_{mn}^{(1,l)} = Pr [ \mathbf{H}_l = \mathbf{h}_n \mid \mathbf{H}_1 = \mathbf{h}_m ].$$

Equation (27) can be easily written in a matrix form as

$$K \text{Var}_{P\gamma}[\hat{Z}_{MN,P\gamma}(K)] = (\underline{Q}^{\gamma})^t \left[ \mathbf{P}^{\gamma} - \mathbf{P}^{\gamma} \mathbf{A} + 2 \mathbf{P}^{\gamma} \sum_{l=1}^{K-1} \frac{K-l}{K} (\mathbf{P}^l - \mathbf{A}) \right] \underline{Q}^{\gamma}, \quad (28)$$

where all the matrices are  $R^{MN} \times R^{MN}$  matrices. Matrix  $\mathbf{P}^{\gamma}$  is a diagonal matrix with elements the  $R^{MN}$  probabilities  $P_{MN}^{\gamma}(\mathbf{h})$ ,  $\mathbf{h} \in E_H^{MN}$ . Matrix  $\mathbf{A}$  has  $R^{MN}$  identical rows which are equal to the diagonal of  $\mathbf{P}^{\gamma}$ , whereas,  $\mathbf{P}^l$  is the  $l$  power of the transition probability matrix  $\mathbf{P}$ . Finally,  $\underline{Q}^{\gamma}$  is an  $R^{MN}$ -dimensional vector, with elements  $Q_{MN,P\gamma}(\mathbf{h})$ ,  $\mathbf{h} \in E_H^{MN}$ . For an ergodic transition probability matrix  $\mathbf{P}$  the *fundamental matrix*  $\mathbf{F}$ , given by  $\mathbf{F} = (\mathbf{I} - \mathbf{P} + \mathbf{A})^{-1}$ , exists, and [45]

$$\mathbf{F} = \mathbf{I} + \lim_{K \rightarrow +\infty} \sum_{l=1}^{K-1} \frac{K-l}{K} (\mathbf{P}^l - \mathbf{A}). \quad (29)$$

From eqs. (28) and (29) we obtain

$$\lim_{K \rightarrow +\infty} K \text{Var}_{P^\gamma} [\hat{Z}_{MN, P^\gamma}(K)] = (\underline{Q}^\gamma)^t \left[ 2 \mathbf{P}^\gamma \mathbf{F} - \mathbf{P}^\gamma - \mathbf{P}^\gamma \mathbf{A} \right] \underline{Q}^\gamma,$$

or

$$\lim_{K \rightarrow +\infty} \text{Var}_{P^\gamma} [\hat{Z}_{MN, P^\gamma}(K)] = 0. \quad (30)$$

From eq. (30) and the *Tchebycheff inequality* we obtain eq. (24); therefore, *Algorithm II* results<sup>4</sup> in an unbiased and consistent estimator for  $Z_{MN}$ .

We would also like to get a practical idea of the accuracy of the resulting estimate. However, since the samples  $\mathbf{h}_k$  are no longer independent, we cannot use the sample variance given by eq. (25). In the case of *Algorithm II*, eq. (25) will most probably underestimate the variance, since it does not take into consideration the correlations among consecutive realizations. To overcome this difficulty, we may use the ideas in [48].

In many instances  $Z_{MN} \rightarrow +\infty$  with increasing lattice size (i.e., as  $M, N \rightarrow +\infty$ ). Additionally, for many realizations  $\mathbf{h}$ ,  $P_{MN}^*(\mathbf{h}) \rightarrow 0$ , whereas,  $A_{MN}(\mathbf{h}) \rightarrow +\infty$  as  $M, N \rightarrow +\infty$ . Therefore, eq. (5c), or equivalently, eq. (11), may be violated. Since most applications require use of rather large lattices, we have to examine the problems introduced by this behavior and modify our Monte-Carlo calculation procedure. It is clear that, in these cases, a simple partition function calculation algorithm (like *Algorithm I*, or *Algorithm II*) will suffer from *overflow* problems. A first step towards the remedy of this problem is to estimate the quantity

$$f(Z_{MN}) = \frac{1}{MN} \ln(Z_{MN}), \quad (31)$$

<sup>4</sup> The central limit theorem, developed in [46], [47], is directly applicable in our case, and, therefore, our estimator is also asymptotically normal.

instead of  $Z_{MN}$ , by (see also eqs. (8b) and (10b))

$$\hat{f}_{K,P\gamma}(Z_{MN}) = \frac{1}{MN} \ln(\hat{Z}_{MN,P\gamma}(K)) = \ln(Q_{MN,P\gamma}(\mathbf{h}_1)) + \ln\left[\frac{1}{K} \sum_{k=1}^K \frac{Q_{MN,P\gamma}(\mathbf{h}_k)}{Q_{MN,P\gamma}(\mathbf{h}_1)}\right], \quad (32a)$$

with

$$\ln(Q_{MN,P\gamma}(\mathbf{h}_1)) = \sum_{i=1}^M \sum_{j=1}^N \ln\left[q_{ij,P\gamma}(h_{ij}^{(1)}, h_{i-1,j}^{(1)}, h_{i-1,j-1}^{(1)}, h_{i,j-1}^{(1)})\right], \quad (32b)$$

where  $q_{ij,P\gamma}(x, y, z, \omega)$  is given by eq. (20) or (23). This is a reasonable approach for two main reasons: (a)  $f(Z_{MN})$  will not necessarily become infinite as  $M, N, Z_{MN} \rightarrow +\infty$  (see for example the case of the Ising model, discussed in Section VI), or, at least, it will be much smaller than  $Z_{MN}$ ; and, (b) we are usually interested in computing the logarithm of the partition function rather than the partition function itself (e.g., in the case of maximum likelihood parameter estimation). However, computing the summation in eq. (32a) may still result in overflow.

To overcome this problem observe that (see also eqs. (8b), (9) and (31))

$$\begin{aligned} f(Z_{MN}) &= \frac{1}{MN} \ln \left[ \sum_{\text{states } \mathbf{h}} Q_{MN,P\gamma}(\mathbf{h}) P_{MN}^{\gamma}(\mathbf{h}) \right] \\ &= \frac{1}{MN} \ln \left[ \sum_{\text{states } \mathbf{h}} \left[ \prod_{i=1}^M \prod_{j=1}^N \frac{q_{ij,P\gamma}^{(\max)}(h_{ij}, h_{i-1,j}, h_{i-1,j-1}, h_{i,j-1})}{q_{ij,P\gamma}^{(\max)}} \right] P_{MN}^{\gamma}(\mathbf{h}) \right] \\ &= \frac{1}{MN} \ln \left[ \sum_{\text{states } \mathbf{h}} q_{\max,P\gamma}^{MN} Q'_{MN,P\gamma}(\mathbf{h}) P_{MN}^{\gamma}(\mathbf{h}) \right] \\ &= \ln q_{\max,P\gamma} + \frac{1}{MN} \ln \left[ \sum_{\text{states } \mathbf{h}} Q'_{MN,P\gamma}(\mathbf{h}) P_{MN}^{\gamma}(\mathbf{h}) \right], \end{aligned} \quad (33)$$

where

$$q_{ij,P\gamma}^{(\max)} = \max_{(x,y,z,\omega) \in E_H^4} \left[ q_{ij,P\gamma}(x, y, z, \omega) \right] \quad ; \quad q_{\max,P\gamma} = \left[ \prod_{i=1}^M \prod_{j=1}^N q_{ij,P\gamma}^{(\max)} \right]^{\frac{1}{MN}},$$

and

$$Q'_{MN, P\gamma}(\mathbf{h}) = \prod_{i=1}^M \prod_{j=1}^N \frac{q_{ij, P\gamma}(h_{ij}, h_{i-1, j}, h_{i-1, j-1}, h_{i, j-1})}{q_{ij, P\gamma}^{(max)}}.$$

From eq. (33) we see that, in order to compute  $f(Z_{MN})$  we have to calculate the expectation  $E_{P\gamma}[Q'_{MN, P\gamma}(\mathbf{h})]$ . This can be done by a simple modification of the previous two algorithms. Notice that, in this case, Monte-Carlo calculations can proceed with no overflow, even in the case of large lattices, because  $Q'_{MN, P\gamma}(\mathbf{h}) \leq 1 < +\infty$ , for all realizations  $\mathbf{h}$ ; therefore, condition (11) is satisfied. Equation (33) yields the following Monte-Carlo estimate for  $f(Z_{MN})$

$$\hat{f}_{K, P\gamma}(Z_{MN}) = \frac{1}{MN} \ln(\hat{Z}_{MN, P\gamma}(K)) = \ln q_{max, P\gamma} + \frac{1}{MN} \ln \left[ \frac{1}{K} \sum_{k=1}^K Q'_{MN, P\gamma}(\mathbf{h}_k) \right],$$

where  $\mathbf{h}_k$  are samples drawn from the joint probability distribution  $P_{MN}^{\gamma}(\mathbf{h})$ .

It is worthwhile noticing that the estimator  $\ln(\hat{Z}_{MN, P\gamma}(K))$  will only asymptotically (i.e., for  $K \rightarrow +\infty$ ) yield an unbiased estimate for  $\ln(Z_{MN})$  [49]. To see this, and for the case of *Algorithm I*, use the central limit theorem for  $\hat{Z}_{MN, P\gamma}(K)$  and the fact that the function  $\ln(\cdot)$  has a nonzero derivative at  $Z_{MN}$  to prove that  $\ln(\hat{Z}_{MN, P\gamma}(K))$  is asymptotically normal with mean  $\ln(Z_{MN})$  and variance  $1/K (\sigma_{MN, P\gamma}/Z_{MN})^2$ , where

$$\sigma_{MN, P\gamma}^2 = \sum_{states \mathbf{h}} Q_{MN, P\gamma}^2(\mathbf{h}) P_{MN}^{\gamma}(\mathbf{h}) - Z_{MN}^2.$$

A similar result can be proved for the case of *Algorithm II*. Finally, we can easily show that  $1/MN \ln(\hat{Z}_{MN, P\gamma}(K))$  is a consistent estimator for  $1/MN \ln(Z_{MN})$ , when either one of the two algorithms is used.

The practical implementation of the previous ideas requires the development of fast algorithms. There are many ways to improve performance. For example, much time is devoted to generating uniformly distributed random numbers. The development of a good, fast and, probably, machine-dependent

random number generator will definitely result in computational savings. Another problem is the fact that the rejection rate approaches 100% in low temperatures. To avoid "being stuck" at the same realization for a long time, one might try to use some "a-priori" knowledge for the MC-GRF parameters and perform various simulations with different suitable configurations of the state space, averaging the obtained results with appropriate weights. An ingenious idea to overcome a similar problem is the "n-fold algorithm" proposed in [50]. Since in low temperatures most GRF's will favor realizations which are characterized by large clusters, one can try to break these clusters into smaller ones and then assign a random spin value to the whole cluster. This idea is used in [51] to obtain an efficient and fast Monte-Carlo simulation algorithm. It is also important to notice that *Algorithm II* can be directly implemented on the Monte-Carlo simulation machines developed in [52-55]. These are special purpose computers developed for the study of various properties of GRF's. Implementation of our algorithm on such machines may result in a substantial reduction of computational time.

Traditionally, stochastic methods are used to simulate GRF's and calculate various quantities related to them. Our method fits to this framework. However, some researchers have adopted deterministic, or pseudo-deterministic, evolution algorithms, like the *microcanonical simulation algorithm* in [56], [57]. For such algorithms, it is much more difficult to prove convergence and study statistical properties of the resulting estimators. Additionally, it is not clear if they can result in significant computational savings. However, we believe that it is extremely interesting to pursue many of these methods further. We haven't done so here though, since our major objective is to present fundamental ideas, and a simple algorithm, for the calculation of  $Z_{MN}$ .

## V. CALCULATION OF PARTITION FUNCTION DERIVATIVES

So far, we have discussed the problem of computing the partition function  $Z_{MN}$  of a GRF. We have proposed two methods which require sampling from two different MC-GRFs, namely,  $P_{MN}^*$  and  $P_{MN}^{**}$ . *Method 2* requires the computation of the  $R^4$  first-order derivatives of the logarithm of the partition function with respect to the LTF (see eq. (22)); i.e., it requires computation of the vector

$$\nabla_{\underline{\sigma}} \ln Z_{MN} = \left\{ \frac{\partial \ln Z_{MN}}{\partial \sigma(x, y, z, \omega)}, (x, y, z, \omega) \in E_H^4 \right\}. \quad (34)$$

The analytical computation of eq. (34) is not possible in general (except in some special cases). However, this calculation can be approximated via Monte-Carlo simulations by using a scheme similar to that used in statistical mechanics for the calculation of various thermodynamic properties of large-scale systems [58]. In addition to the derivatives in eq. (34) we may also want to compute the *internal energy* (which is related to first-order derivatives of the partition function) and the *specific heat* (which is related to second-order derivatives of the partition function) of a general GRF. These quantities provide valuable means for detecting and studying phase transitions. All these computations can be approximated by a simple Monte-Carlo simulation algorithm, as it will be demonstrated in this section.

Let us study the calculation of the derivatives of the logarithm<sup>5</sup> of the partition function  $Z_{MN}$  with respect to the LTF  $\sigma(x, y, z, \omega)$ . We shall limit our analysis to the computation of first- and second-order derivatives. However, our presentation can be trivially extended to include the computation of any higher-order derivative. In the following, we shall denote by  $v_k(x, y, z, \omega)$  the

---

<sup>5</sup> We prefer to calculate the derivatives of the logarithm of the partition function, instead of the derivatives of the partition function itself, for reasons similar to those explained at the end of Section IV.

total number of elementary squares  $(h_{ij}^{(k)}, h_{i-1,j}^{(k)}, h_{i-1,j-1}^{(k)}, h_{i,j-1}^{(k)}) = (x, y, z, \omega)$  which appear in a realization  $\mathbf{h}_k$  of the GRF  $[H]$ . From eqs. (2b) and (3a) we have that

$$Z_{MN} = \sum_{\text{states } \mathbf{h}_l} \prod_{(q,u,v,w) \in E_H^4} \sigma(q,u,v,w)^{v_l(q,u,v,w)} = \sum_{\text{states } \mathbf{h}_l} A_{MN}(\mathbf{h}_l), \quad (35a)$$

where

$$A_{MN}(\mathbf{h}_l) = \prod_{(q,u,v,w) \in E_H^4} \sigma(q,u,v,w)^{v_l(q,u,v,w)}. \quad (35b)$$

Differentiating eq. (35a) with respect to the LTF  $\sigma(x, y, z, \omega)$  we obtain

$$\begin{aligned} Z_{MN}^{(1)}(x, y, z, \omega) &= \frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial \sigma(x, y, z, \omega)} \\ &= \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega)} \frac{\sum_{\text{states } \mathbf{h}_l} v_l(x, y, z, \omega) A_{MN}(\mathbf{h}_l)}{\sum_{\text{states } \mathbf{h}_l} A_{MN}(\mathbf{h}_l)} \\ &= \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega)} \sum_{\text{states } \mathbf{h}_l} v_l(x, y, z, \omega) \pi_{MN}(\mathbf{h}_l), \end{aligned} \quad (36)$$

whereas, differentiating eq. (36) with respect to  $\sigma(q, u, v, w)$ , we obtain (see also eq. (35))

$$\begin{aligned} Z_{MN}^{(2)}(x, y, z, \omega; q, u, v, w) &= \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial \sigma(x, y, z, \omega) \partial \sigma(q, u, v, w)} \\ &= \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega) \sigma(q, u, v, w)} \times \\ &\times \left\{ \sum_{\text{states } \mathbf{h}_l} \left[ v_l(x, y, z, \omega) v_l(q, u, v, w) - v_l(x, y, z, \omega) \delta_{xyza}^{quvw} \right] \pi_{MN}(\mathbf{h}_l) \right. \\ &\quad \left. - \left[ \sum_{\text{states } \mathbf{h}_l} v_l(q, u, v, w) \pi_{MN}(\mathbf{h}_l) \right] \left[ \sum_{\text{states } \mathbf{h}_l} v_l(x, y, z, \omega) \pi_{MN}(\mathbf{h}_l) \right] \right\}, \end{aligned} \quad (37)$$

where

$$\delta_{xyz\omega}^{quvw} = \begin{cases} 1 & , \text{ if } (x, y, z, \omega) = (q, u, v, w) \\ 0 & , \text{ if } (x, y, z, \omega) \neq (q, u, v, w) \end{cases}$$

The quantities defined by equations (36) and (37) can be efficiently calculated via Monte-Carlo simulations based on drawing samples directly from the Gibbs distribution (2). This is a quite difficult problem in practice, whose solution can only be obtained asymptotically, therefore the resulting Monte-Carlo estimates are biased. In practice, we will generate a Markov Chain of realizations  $\mathbf{h}_k$ ,  $k = 1, 2, \dots, K$ , with suitable transition probabilities, that will asymptotically be distributed according to the Gibbs distribution, i.e.,  $\mathbf{h}_k \underset{k \rightarrow \infty}{\sim} \pi_{MN}$ , or  $\lim_{k \rightarrow \infty} Pr[\mathbf{H} = \mathbf{h}_k] = \pi_{MN}(\mathbf{h}_k)$ , given by eq. (2). In order to reduce the variance of the resulting estimates, we discard the  $K_1$  initial configurations of this Markov Chain, where  $K_1$  is a sufficiently large number, such that  $Pr[\mathbf{H} = \mathbf{h}_k] = \pi_{MN}(\mathbf{h}_k)$ , for  $k > K_1$ . Then, we consider the  $K_2$  next states of the Markov Chain, i.e.,  $\mathbf{h}_k$ ,  $k = K_1+1, K_1+2, \dots, K_1+K_2$ , on which we form ergodic averages. This results in

$$\hat{Z}_{MN}^{(1)}(x, y, z, \omega; K_1, K_2) = \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega)} \frac{1}{K_2} \sum_{k=K_1+1}^{K_1+K_2} v_k(x, y, z, \omega), \quad (38a)$$

and

$$\begin{aligned} \hat{Z}_{MN}^{(2)}(x, y, z, \omega; q, u, v, w; K_1, K_2) &= \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega) \sigma(q, u, v, w)} \times \\ &\times \left\{ \frac{1}{K_2} \sum_{k=K_1+1}^{K_1+K_2} \left[ v_k(x, y, z, \omega) v_k(q, u, v, w) - v_k(x, y, z, \omega) \delta_{xyz\omega}^{quvw} \right] \right. \\ &\left. - \frac{1}{K_2^2} \left[ \sum_{k=K_1+1}^{K_1+K_2} v_k(x, y, z, \omega) \right] \left[ \sum_{k=K_1+1}^{K_1+K_2} v_k(q, u, v, w) \right] \right\}, \quad (38b) \end{aligned}$$

which are the Monte-Carlo estimates of the the first- and second-order derivatives of the logarithm of the partition function with respect to the LTF's.

There are many important applications which require the calculation of the partition function and its derivatives. An anticipated application is the development of an optimal algorithm for the maximum-likelihood estimation of the LTF of a GRF from a given realization. We shall not expand on this subject here; instead, we shall illustrate the use of these calculations for the computational study of phase transitions.

The *internal energy*  $E_{MN}(T)$  and the *specific heat*  $C_{MN}(T)$ , given by

$$E_{MN}(T) = \frac{1}{MN} T^2 \frac{\partial \ln Z_{MN}}{\partial T}, \quad (39a)$$

and

$$C_{MN}(T) = \frac{\partial E_{MN}(T)}{\partial T} = \frac{2}{T} E_{MN}(T) + \frac{1}{MN} T^2 \frac{\partial^2 \ln Z_{MN}}{\partial T^2}, \quad (39b)$$

respectively, are two important thermodynamic quantities, associated with a GRF, which allow the study of phase transitions [5]. If  $T_c$  is a temperature such that

$$\lim_{T \rightarrow T_c} \lim_{M, N \rightarrow +\infty} C_{MN}(T) = +\infty, \quad (40)$$

then we say that the GRF is in *phase transition* at *critical temperature*  $T_c$ . The study of phase transition is very important, being the subject of many disciplines, including statistical mechanics [2], [3-5], [29], [34], [44], [59], probability and information theory [2], [32-34], [44], [60], [61] and image processing [12], [34], [44]. From eqs. (39) and (40) we see that the critical temperature  $T_c$  is the temperature at which the internal energy  $E_{MN}(T)$ , or the first-order derivative  $\partial \ln Z_{MN} / \partial T$ , is discontinuous. Observe that

$$\frac{\partial \ln Z_{MN}}{\partial T} = \sum_{(x, y, z, \omega) \in E_H^4} \frac{\partial \ln Z_{MN}}{\partial \sigma(x, y, z, \omega)} \frac{\partial \sigma(x, y, z, \omega)}{\partial T}, \quad (41)$$

and, therefore, the internal energy  $E_{MN}(T)$  can be approximated by (see also eqs. (36), (38a), (39a) and (41))

$$\begin{aligned} \hat{E}_{MN}(T; K_1, K_2) &= T^2 \sum_{(x,y,z,\omega) \in E_H^4} \hat{Z}_{MN}^{(1)}(x,y,z,\omega; K_1, K_2) \frac{\partial \sigma(x,y,z,\omega)}{\partial T} \\ &= T^2 \frac{1}{MN} \frac{1}{K_2} \sum_{k=K_1+1}^{K_1+K_2} \left[ \sum_{(x,y,z,\omega) \in E_H^4} v_k(x,y,z,\omega) \frac{\partial \ln \sigma(x,y,z,\omega)}{\partial T} \right]. \end{aligned} \quad (42)$$

At the critical temperature, the second-order derivative  $\partial^2 \ln Z_{MN} / \partial T^2$  will necessarily become infinite. From eqs. (36)-(38), (39b) and (42) we can easily show that the specific heat  $C_{MN}(T)$  can be approximated by

$$\begin{aligned} \hat{C}_{MN}(T; K_1, K_2) &= T^2 \sum_{(x,y,z,\omega) \in E_H^4} \left\{ \hat{Z}_{MN}^{(1)}(x,y,z,\omega; K_1, K_2) \left[ \frac{2}{T} \frac{\partial \sigma(x,y,z,\omega)}{\partial T} + \frac{\partial^2 \sigma(x,y,z,\omega)}{\partial T^2} \right] \right. \\ &\quad \left. + \frac{\partial \sigma(x,y,z,\omega)}{\partial T} \left[ \sum_{(q,u,v,w) \in E_H^4} \hat{Z}_{MN}^{(2)}(x,y,z,\omega; q,u,v,w; K_1, K_2) \frac{\partial \sigma(q,u,v,w)}{\partial T} \right] \right\} \\ &= \frac{1}{T^2} \frac{1}{MN} \frac{1}{K_2} \left\{ \sum_{k=K_1+1}^{K_1+K_2} \left[ \sum_{(x,y,z,\omega) \in E_H^4} v_k(x,y,z,\omega) T^2 \frac{\partial \ln \sigma(x,y,z,\omega)}{\partial T} \right]^2 \right. \\ &\quad \left. - \frac{1}{K_2} \left[ \sum_{k=K_1+1}^{K_1+K_2} \sum_{(x,y,z,\omega) \in E_H^4} v_k(x,y,z,\omega) T^2 \frac{\partial \ln \sigma(x,y,z,\omega)}{\partial T} \right]^2 \right\}. \end{aligned} \quad (43)$$

To effectively compute the Monte-Carlo estimates given by eqs. (38a), (42) and (43) we employ the Gibbs Sampler algorithm with a lexicographic site updating, by generating a Markov Chain which asymptotically reaches the Gibbs distribution. The overall procedure is summarized in the following *Algorithm III*.

**ALGORITHM III:***Initialization*

1. Generate, lexicographically, a realization  $\mathbf{h}_1$  of the MC-GRF with probability  $P_{MN}^*(\mathbf{h})$ , given by eq. (7), with  $\tau_{ij}(x, y, z, \omega) = \tau_{ij}^*(x, y, z, \omega)$ , given by eq. (19). Set  $k \leftarrow 1$ .
2. Set  $h = h_k(s_p)$ , with  $p = (k - 1) \text{ modulo } MN + 1$ , where  $s_1, s_2, \dots, s_{MN}$ , is a lexicographic ordering of the sites in  $\Lambda_{MN}$ . Denote site  $s_p$  by  $(i, j)$ .
3. Draw a value  $\phi$  from probability  $d_n / (\sum_{l=1}^R d_l)$ ,  $n = 1, 2, \dots, R$ , where

$$\begin{aligned} d_l = & \sigma(\phi_l, h_{i-1, j}^{(k)}, h_{i-1, j-1}^{(k)}, h_{i, j-1}^{(k)}) \times \sigma(h_{i, j+1}^{(k)}, h_{i-1, j+1}^{(k)}, h_{i-1, j}^{(k)}, \phi_l) \\ & \times \sigma(h_{i+1, j}^{(k)}, \phi_l, h_{i, j-1}^{(k)}, h_{i+1, j-1}^{(k)}) \times \sigma(h_{i+1, j+1}^{(k)}, h_{i, j+1}^{(k)}, \phi_l, h_{i+1, j}^{(k)}), \end{aligned} \quad (44)$$

for  $l = 1, 2, \dots, R$ .

4. Set  $h_{ij}^{(k+1)} = \phi$ ,  $h_{mn}^{(k+1)} = h_{mn}^{(k)}$ , for every  $(m, n) \neq (i, j)$ . Set  $k \leftarrow k + 1$ .
5. If  $k = K_1 + 2$  compute  $v_{K_1+1}(x, y, z, \omega)$ , for all  $(x, y, z, \omega) \in E_H^4$ , and

$$TERM_{K_1+1} = T^2 \sum_{(x, y, z, \omega) \in E_H^4} v_{K_1+1}(x, y, z, \omega) \frac{\partial \ln \sigma(x, y, z, \omega)}{\partial T}.$$

Then,

$$der_{K_1+1}(x, y, z, \omega) = v_{K_1+1}(x, y, z, \omega), \text{ for every } (x, y, z, \omega) \in E_H^4,$$

$$SUM 1_{K_1+1} = TERM_{K_1+1}, \quad SUM 2_{K_1+1} = TERM_{K_1+1}^2;$$

otherwise, go to step 2.

*Main Iteration*

6. Set  $h = h_k(s_p)$ , with  $p = (k - 1) \text{ modulo } MN + 1$ , where  $s_1, s_2, \dots, s_{MN}$ , is a lexicographic ordering of the sites in  $\Lambda_{MN}$ . Denote site  $s_p$  by  $(i, j)$ .
7. Draw a value  $\phi$  from probability  $d_n / (\sum_{l=1}^R d_l)$ ,  $n = 1, 2, \dots, R$ , where  $d_l$ ,  $l = 1, 2, \dots, R$  are given by eq. (44).

8. If  $\phi = h_{ij}^{(k)}$ , go to step 9; otherwise, compute  $v_{k+1}(x, y, z, \omega)$  of the new realization, for all  $(x, y, z, \omega) \in E_H^4$ . Also, compute:

$$TERM_{k+1} = T^2 \sum_{(x, y, z, \omega) \in E_H^4} v_{k+1}(x, y, z, \omega) \frac{\partial \ln \sigma(x, y, z, \omega)}{\partial T}.$$

9. Set

$$der_{k+1}(x, y, z, \omega) \leftarrow der_k(x, y, z, \omega) + v_{k+1}(x, y, z, \omega), \text{ for every } (x, y, z, \omega) \in E_H^4,$$

$$SUM 1_{k+1} \leftarrow SUM 1_k + TERM_{k+1}, \quad SUM 2_{k+1} \leftarrow SUM 2_k + TERM_{k+1}^2.$$

10. If  $k + 1 = K_1 + K_2$ , set

$$\hat{Z}_{MN}^{(1)}(x, y, z, \omega; K_1, K_2) = \frac{1}{MN} \frac{1}{\sigma(x, y, z, \omega)} \frac{1}{K_2} der_{K_1+K_2}(x, y, z, \omega),$$

$$\hat{E}_{MN}(T; K_1, K_2) = \frac{1}{MN} \frac{1}{K_2} SUM 1_{K_1+K_2},$$

$$\hat{C}_{MN}(T; K_1, K_2) = \frac{1}{T^2} \frac{1}{MN} \frac{1}{K_2} \left[ SUM 2_{K_1+K_2} - \frac{1}{K_2} SUM 1_{K_1+K_2}^2 \right],$$

and stop; otherwise, set  $h_{ij}^{(k+1)} = \phi$ ,  $h_{mn}^{(k+1)} = h_{mn}^{(k)}$ , for every  $(m, n) \neq (i, j)$ ,  $k \leftarrow k + 1$ , and go to step 6.

The obtained estimates will be consistent, but only asymptotically unbiased.

## VI. SIMULATION EXPERIMENTS

We have demonstrated the fact that the partition function of a general GRF can be calculated by using the Monte-Carlo schemes discussed in Section IV. Two major questions have to be answered at this point. The first is how well the proposed algorithms work. This question is essential, because, in practice, there are many factors that can prevent us from getting the expected results. Some of these factors are: (a) the quality of the random number generator used in the program, (b) the truncation errors incorporated in the

computation of various sums; (c) the number  $K$  of samples used for the computation of the estimates; and, (d) the numbers  $K_1$  and  $K_2$  used in *Method 2* for the calculation of the first-order derivatives of the partition function (see Section V). The mishandling of some, or all, of these factors may give misleading results. The second question concerns the relative merits of *Method 1* and *Method 2*. We are interested in knowing which method gives more accurate results, and which is more computationally efficient. Additionally, we would like to demonstrate the fact that sampling from the probability distribution of a MC-GRF with the i.i.d. choice for the LTF fails to give reasonable estimates, whereas, our methods constitute a vast improvement over the i.i.d. case. All these can be accomplished by comparing our computational results with analytical results obtained by either performing the required summations over all  $R^{MN}$  possible realizations of the GRF  $[H]$ , or by using analytically known solutions. The first approach is limited to GRF's defined over small lattices, whereas, the second approach is usually limited to the *2-D Ising model* with no external magnetic field and nearest neighbor interactions [2], [5], [32-34], [59], [61-66]. We shall present various comparisons, by employing both approaches, and demonstrate that the proposed methods provide highly accurate results.

In the first set of experiments we consider five binary GRF's, at different temperatures, with a second-order neighborhood system and a homogeneous LTF, defined initially over a small rectangular lattice of  $4 \times 4$  sites. These GRF's are fully described in terms of 16 parameters  $\underline{\sigma} = \{ \sigma(x, y, z, \omega), x, y, z, \omega = 0, 1 \}$ . A program has been written which estimates the partition function by using *Algorithm II* and by sampling from the three different probability distributions  $P_{MN}^{iid}$ ,  $P_{MN}^*$ , and  $P_{MN}^{**}$ . Estimates of  $\nabla_{\underline{\sigma}} \ln Z_{MN}$ ,  $E_{MN}(T)$  and  $C_{MN}(T)$  are also obtained by using *Algorithm III*. All estimated quantities are compared with the actual ones, which are calculated by performing summations over all possible 64,000 states. Various simulation experiments have been performed for the

three different sampling probability distributions and for a wide range of temperatures. The results are extremely accurate and demonstrate the fact that both our methods work well, converging to the correct values within a small number of iterations (usually less than a thousand). The method based on the i.i.d. choice fails, especially at low temperatures. The number of iterations greatly depends on the choice of the LTF of the underlying sampling probability distribution, as well as on the temperature. At high temperatures, our methods converge extremely fast, whereas, at low temperatures many iterations are necessary. This is a predictable behavior, since, at high temperatures, the GRF and the MC-GRF are both "close" to an i.i.d. random field, whereas, at lower temperatures, the MC-GRF is only an approximation of the original GRF [38]. This is especially obvious when the GRF exhibits strong diagonal interactions between sites  $(i, j)$  and  $(i-1, j+1)$ , in which case, *Method 1* does not perform much better than the i.i.d. case; however, *Method 2* is always superior. This can be seen by calculating the error variances of the different sampling schemes by performing the necessary summations over all the 64,000 states, and by assuming that *Algorithm 1* is used. These calculations show that *Method 2* results in the lowest variance, whereas, in most cases, *Method 1* has lower variance than the i.i.d. case.

Figure 1 depicts realizations of the five GRF's considered here. Each of the five rows corresponds to realizations of one  $128 \times 128$  site GRF at five different temperatures. Phase transition is apparent in these realizations. As the temperature drops from high ( $T > T_c$ ; realizations look random) to low ( $T < T_c$ ; realizations look ordered and structured), the qualitative behavior of the random field changes, and short range interactions among sites develop into long range ones. Table I depicts the different temperatures used for the realizations of Fig. 1, whereas, Table II depicts the values of vector  $\underline{\sigma}$ . In Figs. 2A-2E a comparison of the exact value of  $f(Z_{4,4})$  and the estimated one, obtained by

using *Algorithm II* with all three sampling methods (i.i.d., *Method 1*, and *Method 2*), is depicted. In this case  $K = K_1 = K_2 = 1,600$ . In many instances, the i.i.d. case fails to give a reasonable result, whereas, our methods give perfect results. The exact internal energy  $E_{4,4}(T)$  and specific heat  $C_{4,4}(T)$  are also depicted, as well as their corresponding estimates, obtained by using *Algorithm III*. A high degree of accuracy is achieved. Figures 3A-3E illustrate the behavior of our methods as compared to the i.i.d. case. The minimum number of iterations  $K_p$  necessary to achieve a certain degree of confidence in the quality of our estimates (a 5% confidence interval), is plotted, in a logarithmic scale, as a function of temperature  $T$ , for the five different GRF's of Table II. If we assume an i.i.d. sampling scheme (like that of *Algorithm I*) and that an asymptotically normal distribution is achieved, then we can easily show that

$$K_p = \text{int} \left[ 100.0 \frac{\text{Var}_P(Q_{MN,P})}{Z_{MN}^2} \right] \Rightarrow \text{Pr}[\hat{Z}_{MN,P}(K) \in (0.8Z_{MN}, 1.2Z_{MN})] \geq 0.95,$$

for  $K \geq K_p$ . *Method 2* results in substantial improvement (with respect to the minimum number of iterations) over the i.i.d. case.

From the first set of experiments, we can conjecture that *Method 2* is superior to *Method 1*. Now, we would like to see whether this is true as the lattice size grows. In the sequel, we consider the same five GRF's on larger lattices (up to  $128 \times 128$  sites). We use *Algorithms II* and *III* to obtain estimates of  $Z_{MN}(K)$  (sampling from  $P_{MN}^{\text{iid}}$ ,  $P_{MN}^*$  and  $P_{MN}^{**}$ ), of  $E_{MN}(T)$ , and  $C_{MN}(T)$ . Although exact analytical results are unavailable for general GRF's, we can draw some conclusions about the efficiency of our two methods by checking the behavior of the estimated curves of  $f(\hat{Z}_{MN,P})$  versus the temperature  $T$  as the lattice size becomes larger. Such curves are depicted in Figs. 4A-4E (for  $16 \times 16$  sites), Figs. 5A-5E (for  $32 \times 32$  sites), Figs. 6A-6E (for  $64 \times 64$  sites), and Figs. 7A-7E (for  $128 \times 128$  sites), for the five GRF models of Table II. In all cases  $K = K_1 = K_2 = 10,000MN$ . The shape of these curves indicates that *Method 1*

breaks down at low temperatures (for  $T < T_c$ ), being comparable to *Method 2* at high temperatures (for  $T > T_c$ ). However, it is impossible to judge the accuracy of *Method 2*, especially at low temperatures and at temperatures around the critical temperature, since we have no exact results to compare with. It seems though, that *Method 2* gives reasonably good estimates, except in the case when the random field exhibits strong diagonal interactions between sites  $(i, j)$  and  $(i-1, j+1)$ . This is evident from Figs. 1, 4E, 5E, 6E, and 7E; the estimated curves of the partition function versus  $T$ , obtained from our simulations, are not "smooth enough" at low temperatures, implying large error variance. Notice however that, the remarkably good results obtained from the second set of our experiments, described next, are very encouraging and strongly favor *Method 2* over *Method 1*. Finally, to obtain a rough idea about the location of the critical temperature  $T_c$  of the GRF's under consideration, we plot the internal energy and specific heat curves of the GRF models, for different lattice sizes ( $8 \times 8$ ,  $16 \times 16$ , and  $32 \times 32$ ). The resulting graphs are depicted in Figs. 8A-8E; they allow us to conclude that the break down point of *Method 1* is indeed very close to  $T_c$ . The critical temperature corresponds to the peak of the specific heat curve as the lattice size grows (see eq. (40)).

In the second set of our experiments, we compare our results with analytically known solutions. We consider a special case of a GRF, the two-dimensional Ising model with no external magnetic field and nearest neighbor interactions, defined on a rectangular lattice. This model has been first introduced by the German physicist Ernst Ising in his attempt to statistically formulate the phenomenon of ferromagnetism. Although the Ising model has a very simple LTF, it has been enjoying considerable attention, because it is one of the rare non-trivial (i.e., it exhibits phase transition) GRF models whose partition function and its derivatives can be computed analytically. The neighborhood system for such a model is the first-order system  $N_{ij}^{(1)}$ , defined in Section

II, whereas,  $R = 2$ ,  $E_H = \{-1, +1\}$  and<sup>6</sup>

$$\sigma(X, Y, Z, \Omega) = \exp \left[ \frac{1}{T} (BXY + AX\Omega) \right]. \quad (45)$$

If we assume torodial boundary conditions, we can compute the partition function analytically, as it has been done initially by Onsager [59]. If we define parameters  $\alpha = A/T$  and  $\beta = B/T$ , then the partition function of the Ising model is given by [32], [63] (for all temperatures  $T$  such that  $T \neq T_c$ )

$$Z_{MN} = \frac{1}{2} (2 \sinh 2\alpha)^{\frac{MN}{2}} \left\{ \prod_{r=1}^N [2 \cosh(\frac{M}{2} \gamma_{2r})] + \prod_{r=1}^N [2 \sinh(\frac{M}{2} \gamma_{2r})] \right. \\ \left. + \prod_{r=1}^N [2 \cosh(\frac{M}{2} \gamma_{2r-1})] + \prod_{r=1}^N [2 \sinh(\frac{M}{2} \gamma_{2r-1})] \right\}, \quad (46a)$$

where

$$\cosh \gamma_j = \cosh 2\alpha^* \cosh 2\alpha - \sinh 2\alpha^* \sinh 2\beta \cos(\pi j / N), \quad (46b)$$

for  $j = 1, 2, \dots, 2N$ , and  $\alpha^*$  satisfies

$$\sinh 2\alpha \sinh 2\alpha^* = 1. \quad (46c)$$

Phase transition occurs at the critical temperature  $T_c$  which satisfies the equation

$$\sinh \left[ \frac{2A}{T_c} \right] \sinh \left[ \frac{2B}{T_c} \right] = 1. \quad (47)$$

At the critical temperature  $\gamma_{2N} = 0$ . Following the derivation of eq. (46) in [63] we can easily see that  $\gamma_r > 0$ , for every  $r = 1, 2, \dots, 2N-1$  and that  $\gamma_{2N} > 0$ , if  $T < T_c$ , whereas,  $\gamma_{2N} < 0$ , if  $T > T_c$ . Therefore, eq. (46), together with the identity

<sup>6</sup> We use capital letters for  $x, y, z, \omega$  to denote that they take values in  $\{-1, +1\}$ . Small letters will denote variables  $x, y, z, \omega$  which take values in  $\{0, +1\}$ .

$$\cosh l\theta = 2^{l-1} \prod_{s=0}^{l-1} \left[ \cosh\theta - \cos \frac{(2s+1)\pi}{2l} \right]$$

and the knowledge of the signs of  $\gamma_i$ 's, can be used to calculate the partition function of the Ising model. In the case of large lattices, and for temperatures  $T \neq T_c$ , the following approximation can be used [61]

$$\frac{1}{MN} \ln Z_{MN} \approx \frac{1}{2} \ln(4 \cosh 2\alpha \cosh 2\beta) + B(\alpha, \beta), \quad (48a)$$

where

$$B(\alpha, \beta) = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln \Phi(\tanh 2\alpha, \tanh 2\beta; \theta, \phi) d\theta d\phi, \quad (48b)$$

$$\Phi(x, y; \theta, \phi) = 1 - x(1-y^2)^{\frac{1}{2}} \cos\theta - y(1-x^2)^{\frac{1}{2}} \cos\phi, \quad (48c)$$

with  $x = \tanh 2\alpha$  and  $y = \tanh 2\beta$ . Equation (48b) can be approximated by its Riemann sum; therefore,

$$B(\alpha, \beta) = \frac{1}{2MN} \sum_{j=1}^M \sum_{k=1}^N \ln \Phi(\tanh 2\alpha, \tanh 2\beta; \theta_{jk}, \phi_k), \quad (49a)$$

where

$$\theta_{jk} = \left[ \frac{2j-2}{M} + \frac{2k-1}{MN} \right] \pi, \quad (49b)$$

and

$$\phi_k = \left[ \frac{2k-1}{N} \right] \tau. \quad (49c)$$

Equations (48) and (49) are quite accurate, even in the cases of small lattices; therefore, we shall use them when  $M, N > 20$ . When  $M, N \leq 20$ , eq. (46) should be used.

Equations (48) and (49) can also be used to compute the first- and second-order derivatives of the partition function with respect to  $T$ . Indeed, it is easy to show that,

$$E_{MN}(T) = T^2 \frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial T}, \quad (50a)$$

and

$$C_{MN}(T) = 2T \frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial T} + T^2 \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial T^2}, \quad (50b)$$

where

$$\frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial T} = -\frac{A}{T^2} \frac{1}{MN} \frac{\partial \ln Z_{A,N}}{\partial \alpha} - \frac{B}{T^2} \frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial \beta}, \quad (50c)$$

$$\begin{aligned} \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial T^2} = \frac{1}{MN} \left\{ \frac{2A}{T^3} \frac{\partial \ln Z_{A,N}}{\partial \alpha} + \frac{2B}{T^3} \frac{\partial \ln Z_{MN}}{\partial \beta} + \right. \\ \left. + \frac{A^2}{T^4} \frac{\partial^2 \ln Z_{MN}}{\partial \alpha^2} + \frac{2AB}{T^4} \frac{\partial \ln Z_{MN}}{\partial \alpha \partial \beta} + \frac{B^2}{T^4} \frac{\partial^2 \ln Z_{MN}}{\partial \beta^2} \right\}, \quad (50d) \end{aligned}$$

and, for  $T \neq T_c$ ,

$$\frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial \alpha} = \tanh^2 \alpha + \frac{1}{2MN} \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_\alpha(x, y; \theta_{jk}, \phi_k)}{\Phi(x, y; \theta_{jk}, \phi_k)}, \quad (51a)$$

$$\frac{1}{MN} \frac{\partial \ln Z_{MN}}{\partial \beta} = \tanh^2 \beta + \frac{1}{2MN} \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_\beta(x, y; \theta_{jk}, \phi_k)}{\Phi(x, y; \theta_{jk}, \phi_k)}, \quad (51b)$$

$$\begin{aligned} \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial \alpha^2} = 2(1 - \tanh^2 2\alpha) + \frac{1}{2MN} \left[ \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_{\alpha\alpha}(x, y; \theta_{jk}, \phi_k)}{\Phi(x, y; \theta_{jk}, \phi_k)} \right. \\ \left. - \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_\alpha^2(x, y; \theta_{jk}, \phi_k)}{\Phi^2(x, y; \theta_{jk}, \phi_k)} \right], \quad (51c) \end{aligned}$$

$$\begin{aligned} \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial \beta^2} = 2(1 - \tanh^2 2\beta) + \frac{1}{2MN} \left[ \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_{\beta\beta}(x, y; \theta_{jk}, \phi_k)}{\Phi(x, y; \theta_{jk}, \phi_k)} \right. \\ \left. - \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_\beta^2(x, y; \theta_{jk}, \phi_k)}{\Phi^2(x, y; \theta_{jk}, \phi_k)} \right], \quad (51d) \end{aligned}$$

$$\begin{aligned} \frac{1}{MN} \frac{\partial^2 \ln Z_{MN}}{\partial \alpha \partial \beta} = \frac{1}{2MN} \left[ \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_{\alpha\beta}(x, y; \theta_{jk}, \phi_k)}{\Phi(x, y; \theta_{jk}, \phi_k)} \right. \\ \left. - \sum_{j=1}^M \sum_{k=1}^N \frac{\Phi_\alpha(x, y; \theta_{jk}, \phi_k) \Phi_\beta(x, y; \theta_{jk}, \phi_k)}{\Phi^2(x, y; \theta_{jk}, \phi_k)} \right]. \quad (51e) \end{aligned}$$

In eqs. (51)  $\Phi_\alpha$ ,  $\Phi_{\alpha\alpha}$ ,  $\Phi_\beta$ ,  $\Phi_{\beta\beta}$  and  $\Phi_{\alpha\beta}$  are the first- and second-order partial derivatives of  $\Phi$  with respect to  $\alpha$  and  $\beta$ , given by

$$\Phi_\alpha(x, y; \theta, \phi) = -2(1-x^2)(1-y^2)^{\frac{1}{2}} \cos\theta + 2xy(1-x^2)^{\frac{1}{2}} \cos\phi, \quad (52a)$$

$$\Phi_\beta(x, y; \theta, \phi) = -2(1-y^2)(1-x^2)^{\frac{1}{2}} \cos\theta + 2xy(1-y^2)^{\frac{1}{2}} \cos\phi, \quad (52b)$$

$$\Phi_{\alpha\alpha}(x, y; \theta, \phi) = 8x(1-x^2)(1-y^2)^{\frac{1}{2}} \cos\theta + 4(1-2x^2)y(1-x^2)^{\frac{1}{2}} \cos\phi, \quad (52c)$$

$$\Phi_{\beta\beta}(x, y; \theta, \phi) = 8y(1-y^2)(1-x^2)^{\frac{1}{2}} \cos\theta + 4(1-2y^2)x(1-y^2)^{\frac{1}{2}} \cos\phi, \quad (52d)$$

and

$$\Phi_{\alpha\beta}(x, y; \theta, \phi) = 4y(1-x^2)(1-y^2)^{\frac{1}{2}} \cos\theta + 4x(1-x^2)^{\frac{1}{2}}(1-y^2) \cos\phi. \quad (52e)$$

Equations (48)-(52) are used for the analytical computation of the partition function and its derivatives. To obtain the Monte-Carlo estimates, we have to modify our program, since the random variable  $H(i, j)$  assumes values  $-1, +1$  instead of values  $0, +1$ , and since we have assumed toroidal boundary conditions, instead of free boundary conditions. As the lattice becomes large the effects of the boundary conditions become negligible [2].<sup>7</sup> We can now use the original Monte-Carlo simulation program, used in the first class of experiments, by employing the following transformation

$$X = 2x - 1, \quad Y = 2y - 1, \quad \Omega = 2\omega - 1, \quad (53)$$

which yields binary 0-1 random variables. Substituting eq. (53) into eq. (45) we obtain

$$\sigma(x, y, z, \omega) = \exp \left[ \frac{1}{T} \left[ -(A+B) - (2A+2B)x - 2B y - 2A \omega + 4B xy + 4A x \omega \right] \right].$$

<sup>7</sup> This may not be the case when the temperature is close to the critical temperature, in which case a discrepancy between the analytical and computational result is expected.

In our simulation experiments we have considered four different Ising models defined on various size lattices, up to  $128 \times 128$  sites. These models are viewed as one-parameter models in which  $A$ ,  $B$  are kept fixed and  $T$  varies. Realizations of these GRF's on a  $128 \times 128$  site lattice, at different temperatures, are depicted in Fig. 9. The corresponding values of  $A$  and  $B$ , together with some other useful information, are depicted in Table III. Our objective now is to compare the exact values of the partition function with the estimated ones, obtained by using *Method 1* and *Method 2*. Figures 10A and 10B show the typical convergence behavior of our two methods for the case of the Ising 1 model considered on a  $16 \times 16$  site lattice, at a high temperature (weak couplings) and at a low temperature (strong couplings). The obtained results verify our previously observation that *Method 2* is more accurate at temperatures below the critical temperature than *Method 1*. In Figs. 11A-11D we compare the estimates obtained by using *Method 1* and *Method 2* with the exact results for the four Ising models of Table III considered on a  $32 \times 32$  site lattice. The same quantities are depicted in Figs. 12A-12D, for a  $64 \times 64$  site lattice, and in Figs. 13A-13D for a  $128 \times 128$  site lattice. In all cases  $K = K_1 = K_2 = 10,000 MN$ . The obtained results indicate that *Method 2* provides really good estimates of the partition function, whereas, *Method 1* fails to do so at low temperatures. In Figs. 14A-14D we compare  $\hat{E}_{32,32}(T; K_1, K_2)$  with  $E_{32,32}(T)$ , and  $\hat{C}_{32,32}(T; K_1, K_2)$  with  $C_{32,32}(T)$ , in order to check on the accuracy of *Algorithm III*. We have used  $K_1 = K_2 = 32^2 \times 10,000$  iterations. Finally, we consider the Ising model as a two-parameter model, with respect to the parameters ( $\alpha = A/T$ ,  $\beta = B/T$ ), on a  $16 \times 16$  site lattice. We took  $0.2 < \alpha, \beta < 0.8$ , which corresponds to models with both strong and weak couplings, as it is indicated in Fig. 15. The partition function of the Ising model has been calculated by using *Method 1* (see Fig. 16B) and *Method 2* (see Fig. 16C). The exact partition function is depicted in Fig. 16A. The resulting computational error is shown in Figs. 17A and 17B. These

results clearly illustrate the superiority of *Method 2*.

To conclude, we would like to point out that, although *Method 2* is usually superior to *Method 1* at virtually all temperatures, we prefer to use *Method 1* at high temperatures (i.e., for  $T > T_c$ ) because of its reasonable accuracy and computational efficiency (recall that the implementation of *Method 2* requires the computation of the first-order derivatives of the partition function, with respect to  $\sigma(x, y, z, \omega)$ , which is a quite expensive computation).

## VII. CONCLUSIONS

We have presented a new technique for the estimation of the partition function of a general GRF which is based on approximating a general GRF by a MC-GRF. We have discussed an optimal choice for the MC-GRF in terms of achieving error variance reduction, and we have adopted two reasonable suboptimal choices that have the advantage of simplicity and computability. These choices contain substantial information about the initial GRF, and allow us to achieve significant error variance reduction. The second choice is optimal in terms of minimizing an entropy distance from the given Gibbs distribution. We have also considered different Monte-Carlo algorithms and concluded that the choice of the Gibbs Sampler with lexicographic site updating was the most appropriate. Our proposed techniques result in unbiased and consistent estimates of the partition function. This is a major improvement over existing Metropolis-like simulation algorithms, which sample directly from a distribution that asymptotically approaches the Gibbs distribution, and are capable of estimating only the derivatives of the logarithm of the partition function, and not the logarithm of the partition function itself.

Our two choices of MC-GRF's have resulted in two different methods for the calculation of the partition function. We have carried out many simulation

experiments in order to test the reliability, accuracy and the relative merits of the two approaches. The obtained results have been compared to analytical ones, wherever possible. Our experiments gave remarkably accurate results, demonstrating the fact that we can obtain good estimates within reasonable computational time, for a wide variety of GRF models. We have also examined the behavior of the methods as the lattice size increases and as the temperature approaches the critical temperature, and conjectured that *Method 2* is more reliable than *Method 1* around and below the critical region of the GRF. Nevertheless, we prefer to use *Method 1* at high temperatures. We believe that our methods can be effectively used for the optimal parameter estimation of a general GRF. We would finally like to notice that, the use of computationally efficient algorithms can boost the performance of our methods, and yield even more reliable estimates.

#### ACKNOWLEDGMENTS

This work was supported by the office of Naval Research, Mathematical Sciences Division, under ONR Grant N00014-90-J-1345.

## REFERENCES

- [1] C.J. Preston, *Gibbs States on Countable Sets*. London, England: Cambridge University Press, 1974.
- [2] R. Kindermann and J.L. Snell, *Markov Random Fields and Their Applications*. Contemporary Mathematics, Vol. 1, Providence, Rhode Island: American Mathematical Society, 1980.
- [3] D. Ruelle, *Statistical Mechanics: Rigorous Results*. New York City, New York: Benjamin, 1969.
- [4] C.J. Thomson, *Mathematical Statistical Mechanics*. Princeton, New Jersey: Princeton University Press, 1972.
- [5] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*. London, England: Academic Press, 1982.
- [6] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 192-236, 1974.
- [7] W. Weidlich, "The statistical description of polarization phenomena in society," *British Journal of Mathematical Statistical Psychology*, vol. 24, pp. 251-266, 1971.
- [8] T.R. Welberry and R. Galbraith, "A two-dimensional model of crystal-growth disorder," *Journal of Applied Crystallography*, vol. 6, pp. 87-96, 1973.
- [9] I.G. Enting, "Crystal growth models and Ising models: Disorder points," *Journal of Physics C: Solid State Physics*, vol. 10, pp. 1379-1388, 1977.
- [10] F.S. Cohen, "Markov random fields for image modeling and analysis." In *Modeling and Application of Stochastic Processes*, U.B. Desai, Ed., pp. 243-272. Boston, Massachusetts: Kluwer Academic Publishers, 1986.

- [11] G.R. Cross and A.K. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 25-39, 1983.
- [12] R.C. Dubes and A.K. Jain, "Random field models in image analysis," *Journal of Applied Statistics*, vol. 16, pp. 131-164, 1989.
- [13] M. Hassner and J. Sklansky, "The use of Markov random fields as models of texture," *Computer Graphics and Image Processing*, vol. 12, pp. 357-370, 1980.
- [14] F.R. Hansen and H. Elliott, "Image segmentation using simple Markov field models," *Computer Graphics and Image Processing*, vol. 20, pp. 101-132, 1982.
- [15] H. Derin, H. Elliott, R. Cristi and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 707-720, 1984.
- [16] H. Derin and W.S. Cole, "Segmentation of textured images using Gibbs random fields," *Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 72-98, 1986.
- [17] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 39-55, 1987.
- [18] P.A. Kelly, H. Derin and K.D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1628-1641, 1988.
- [19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

- [20] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, pp. 259-302, 1986.
- [21] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 194-202, 1989.
- [22] M. Kanefsky and C-B. Fong, "Predictive source coding techniques using maximum likelihood prediction for compression of digitized images," *IEEE Transactions on Information Theory*, vol. 30, pp. 722-727, 1984.
- [23] J. Marroquin, S. Mitter and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association*, vol. 82, pp. 76-89, 1987.
- [24] J.W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 458-465, San Diego, California, June 4-8, 1989.
- [25] D.M. Burley, "Closed form approximations for lattice systems." In *Phase Transitions and Critical Phenomena*, Vol. 2, C. Domb and M.S. Green, Eds., pp. 329-374. London, England: Academic Press, 1972.
- [26] R. Kikuchi, "A theory of cooperative phenomena," *Physical Review*, vol. 81, pp. 988-1003, 1951.
- [27] M.F. Sykes, J.W. Essam and D.S. Gaunt, "Derivation of low-temperature expansions for the Ising model of a ferromagnet and an antiferromagnet," *Journal of Mathematical Physics*, vol. 6, pp. 283-298, 1965.
- [28] M.F. Sykes, D.S. Gaunt, J.W. Essam and C.J. Elliott, "Derivation of low-temperature expansions for Ising model VI. Three dimensional lattices - temperature grouping," *Journal of Physics A: Mathematical, Nuclear, and General*, vol. 6, pp. 1507-1516, 1973.

- [29] K.G. Wilson, "The renormalization group and critical phenomena," *Reviews of Modern Physics*, vol. 55, pp. 583-600, 1983.
- [30] B. Gidas, "A renormalization group approach to image processing problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 164-180, 1989.
- [31] J.K. Percus, "Approximation methods in classical statistical mechanics," *Physical Review Letters*, vol. 8, pp. 462-463, 1962.
- [32] D.K. Pickard, "Asymptotic inference for an Ising lattice," *Journal of Applied Probability*, vol. 13, pp. 486-497, 1976.
- [33] D.K. Pickard, "Inference for general Ising models," *Journal of Applied Probability*, vol. 19A, pp. 345-357, 1982.
- [34] D.K. Pickard, "Inference for discrete Markov fields: The simplest non-trivial case," *Journal of the American Statistical Association*, vol. 82, pp. 90-96, 1987.
- [35] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 799-813, 1989.
- [36] L. Younes, "Estimation and annealing for Gibbsian fields," *Annals of the Institute of Henri Poincare*, vol. 24, pp. 269-294, 1988.
- [37] J. Goutsias, "Mutually compatible Gibbs random fields," *IEEE Transactions on Information Theory*, vol. 35, pp. 1233-1249, 1989.
- [38] J. Goutsias, "Unilateral approximation of Gibbs random field images," *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, vol. 53, pp. 240-257, 1991.
- [39] M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods. Volume I: Basics*. New York City, New York: John Wiley and Sons, 1986.

- [40] G.C. Orsak and B. Aazhang, "Constrained solutions in importance sampling via robust statistics," *IEEE Transactions on Information Theory*, vol. 37, pp. 307-316, 1991.
- [41] J. Goutsias, "A theoretical analysis of Monte Carlo algorithms for the simulation of Gibbs random field images," *IEEE Transactions on Information Theory*, vol. 37, No. 6, 1991.
- [42] J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods*. New York City, New York: John Wiley and Sons, 1964.
- [43] J.L. Marroquin, *Probabilistic Solutions of Inverse Problems*. Report LIDS-TH-1500, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. Cambridge, Massachusetts, 1985.
- [44] B. Gidas, "Nonstationary Markov chains and convergence of the annealing algorithm," *Journal of Statistical Physics*, vol. 39, pp. 73-131, 1985.
- [45] J.G. Kemeny and J.L. Snell, *Finite Markov Chains*. Princeton, New Jersey: D. Van Nostrand, 1960.
- [46] R.L. Dobrushin, "Central limit theorem for nonstationary Markov chains I," *Theory of Probability and its Applications*, vol. 1, pp. 65-80, 1956.
- [47] R.L. Dobrushin, "Central limit theorem for nonstationary Markov chains II," *Theory of Probability and its Applications*, vol. 1, pp. 329-383, 1956.
- [48] R. Friedberg and J.E. Cameron, "Test of the Monte Carlo method: Fast simulation of a small Ising lattice," *The Journal of Chemical Physics*, vol. 52, pp. 6049-6058, 1970.
- [49] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. Oakland, California: Holden-Day, 1977.
- [50] A.B. Bortz, M.H. Kalos and J.L. Lebowitz, "A new algorithm for Monte Carlo simulation of Ising spin systems," *Journal of Computational Physics*, vol. 17, pp. 10-18, 1975.

- [51] R.H. Swendsen and J-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letters*, vol. 58, pp. 86-88, 1987.
- [52] R.B. Pearson, J.L. Richardson, and D. Toussaint, "A fast processor for Monte-Carlo simulation," *Journal of Computational Physics*, vol. 51, pp. 241-249, 1983.
- [53] A. Hoogland, J. Spaa, B. Selman, and A. Compagner, "A special-purpose processor for the Monte Carlo simulation of Ising spin systems," *Journal of Computational Physics*, vol. 51, pp. 250-260, 1983.
- [54] A. Margalit, "A parallel algorithm to generate a Markov random field image on a SIMD hypercube machine," *Pattern Recognition Letters*, vol. 9, pp. 263-278, 1989.
- [55] P.D. Hortensius, H.C. Card, R.D. McLeod, and W. Pries, "Importance sampling for Ising computers using one-dimensional cellular automata," *IEEE Transactions on Computers*, vol. 38, pp. 769-774, 1989.
- [56] M. Creutz, "Microcanonical Monte Carlo simulation," *Physical Review Letters*, vol. 50, pp. 1411-1414, 1983.
- [57] G. Bhanot, M. Creutz and H. Neuberger, "Microcanonical simulation of Ising systems," *Nuclear Physics*, vol. B235, pp. 417-434, 1984.
- [58] L.D. Fosdick, "Monte Carlo computation on the Ising lattice." In *Methods of Computational Physics, Vol. 1*, B. Alder, Ed., pp. 245-280. New York City, New York: Academic Press, 1963.
- [59] L. Onsager, "Crystal statistics. I. A two-dimensional model with an order-disorder transition," *Physical Review*, vol. 65, pp. 117-149, 1944.
- [60] T. Berger and Z. Ye, " $\epsilon$ -Entropy and critical distortion of random fields," *IEEE Transactions on Information Theory*, vol. 36, pp. 717-725, 1990.
- [61] D.K. Pickard, "Asymptotic inference for an Ising lattice. II," *Advances in Applied Probability*, vol. 9, pp. 476-501, 1977.

- [62] H.A. Kramers and E.W. Wannier, "Statistics of the two-dimensional ferromagnet. Part II," *Physical Review*, vol. 60, pp. 263-276, 1944.
- [63] B. Kaufman, "Crystal Statistics II. Partition function evaluated by spinor analysis," *Physical Review*, vol. 76, pp. 1232-1243, 1949.
- [64] M. Kac and J.C. Ward, "A combinatorial solution of the two-dimensional Ising model," *Physical Review*, vol. 88, pp. 1332-1337, 1952.
- [65] G.F. Newell and E.W. Montroll, "On the theory of the Ising model of ferromagnetism," *Reviews of Modern Physics*, vol. 25, pp. 353-389, 1953.
- [66] C.A. Hurst and H.S. Green, "New solution of the Ising problem for a rectangular lattice," *The Journal of Chemical Physics*, vol. 33, pp. 1059-1062, 1960.

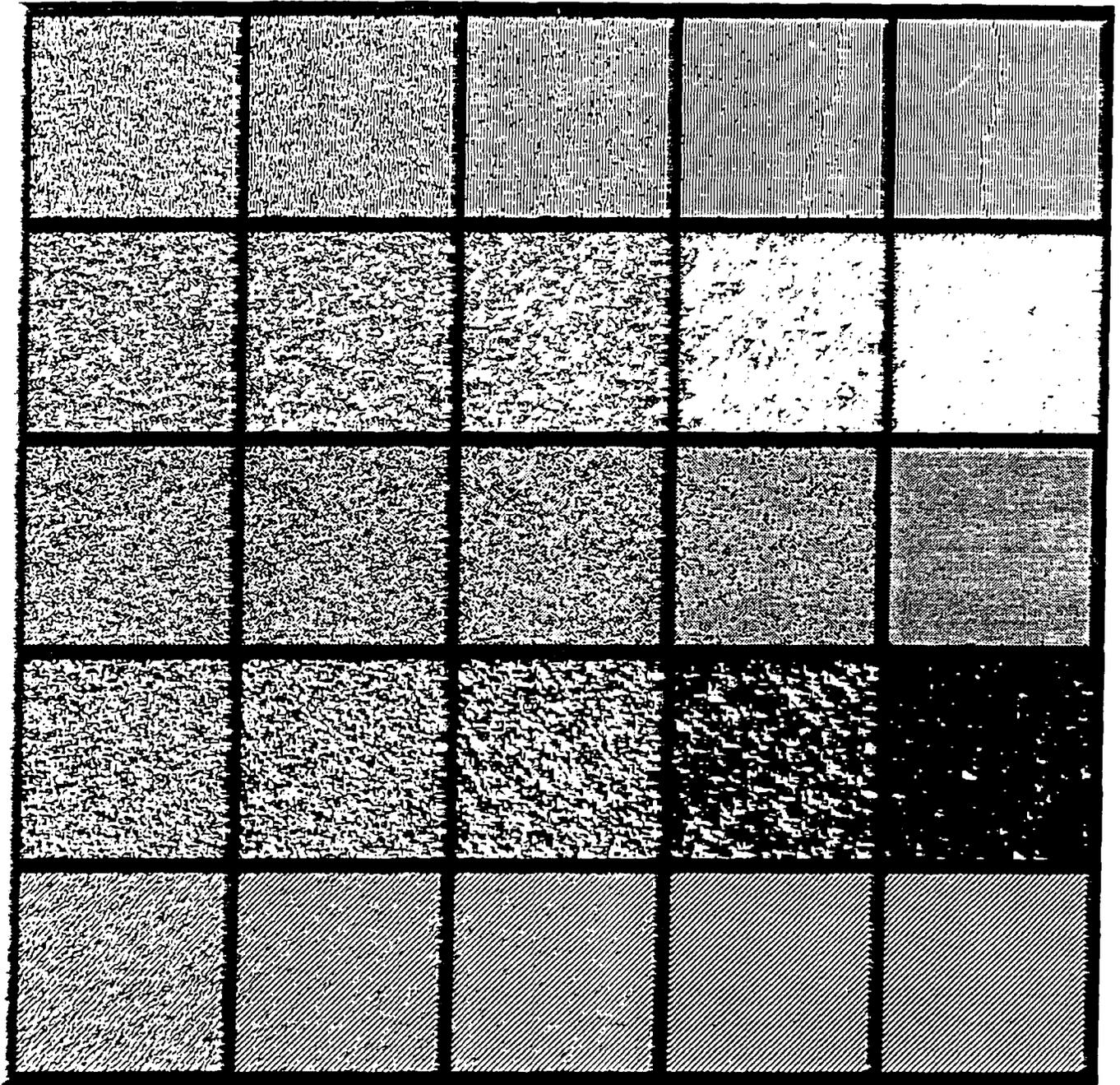


Figure 1: Realizations of five  $128 \times 128$  site GRF's at different temperatures. Each row depicts a GRF model (top to bottom: GRF1, GRF2, GRF3, GRF4, GRF5) at five different temperatures (left to right:  $T_1, T_2, T_3, T_4, T_5$ ). See also Tables I, II.

**Table I:** Temperatures used in the GRF realizations depicted in Fig.1.

Reference name	Temperatures				
	$T_0$	$T_1$	$T_2$	$T_3$	$T_4$
GRF1	2.00	1.30	1.00	0.75	0.50
GRF2	2.00	1.70	1.40	1.20	1.00
GRF3	3.00	2.00	1.30	1.00	0.50
GRF4	3.00	2.00	1.30	1.00	0.80
GRF5	3.00	2.30	2.00	1.30	1.00

**Table II:** LTF's of the five GRF's depicted in Fig.1.

LTF's	GRF1	GRF2	GRF3	GRF4	GRF5
$T \times \ln \sigma(0,0,0,0)$	0.00000	0.00000	0.00000	0.00000	0.00000
$T \times \ln \sigma(0,0,0,1)$	-0.50000	0.00000	0.00000	-0.40000	0.00000
$T \times \ln \sigma(0,0,1,0)$	-0.50000	0.00000	0.00000	0.03000	0.00000
$T \times \ln \sigma(0,0,1,1)$	-1.00000	0.00000	0.00000	-0.72000	0.00000
$T \times \ln \sigma(0,1,0,0)$	1.00000	0.00000	0.00000	-0.91800	0.00000
$T \times \ln \sigma(0,1,0,1)$	0.51500	-1.25000	0.59500	-2.04300	5.00000
$T \times \ln \sigma(0,1,1,0)$	0.50000	0.00000	0.00000	-0.55000	0.00000
$T \times \ln \sigma(0,1,1,1)$	0.01500	-1.84960	0.59500	-2.85500	5.20000
$T \times \ln \sigma(1,0,0,0)$	-0.26000	0.89540	0.00000	-3.20000	0.20000
$T \times \ln \sigma(1,0,0,1)$	-2.76000	-0.94740	-1.25000	-0.60000	-1.80000
$T \times \ln \sigma(1,0,1,0)$	-0.63000	1.22990	0.59500	-4.99000	-4.80000
$T \times \ln \sigma(1,0,1,1)$	-3.13000	-0.68070	-0.65500	-0.24000	-6.70000
$T \times \ln \sigma(1,1,0,0)$	2.84000	-0.86980	-1.25000	-0.64260	0.70000
$T \times \ln \sigma(1,1,0,1)$	0.35500	0.96180	-1.90500	-0.20560	4.10000
$T \times \ln \sigma(1,1,1,0)$	2.47000	-0.66600	-0.65500	-0.97210	-4.20000
$T \times \ln \sigma(1,1,1,1)$	-0.01500	0.90200	-1.31000	-0.11510	0.40000

Figure 2A

GRF1, 4 x 4 sites

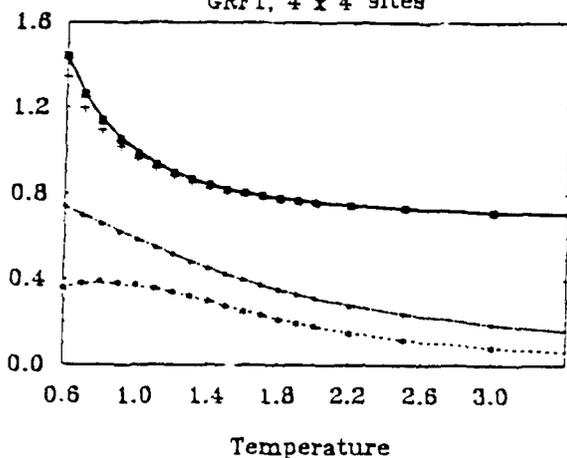


Figure 2B

GRF2, 4 x 4 sites

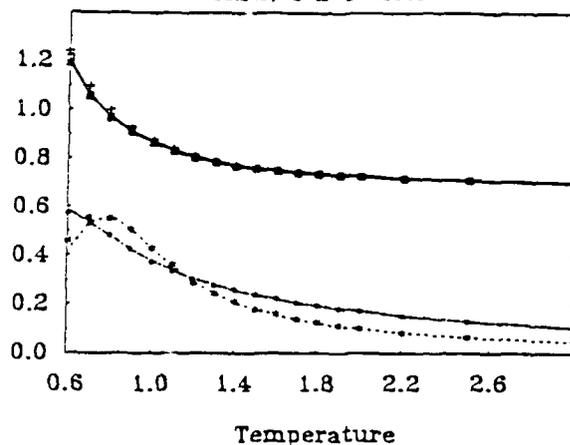


Figure 2C

GRF3, 4 x 4 sites

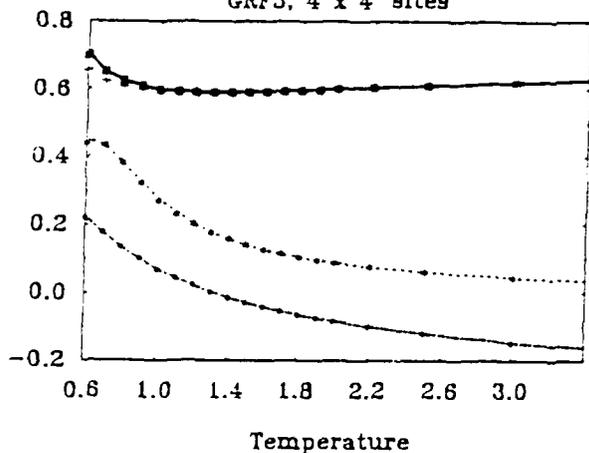


Figure 2D

GRF4, 4 x 4 sites

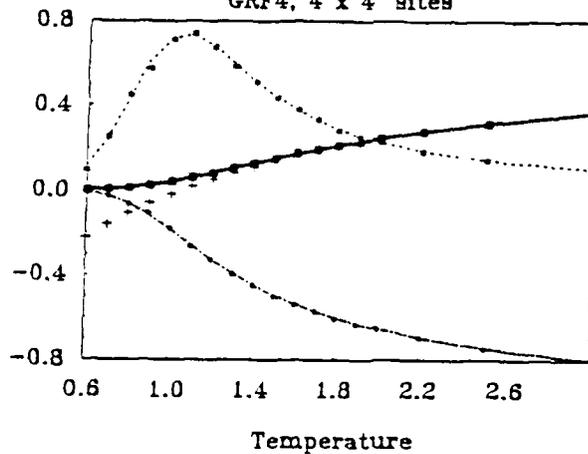
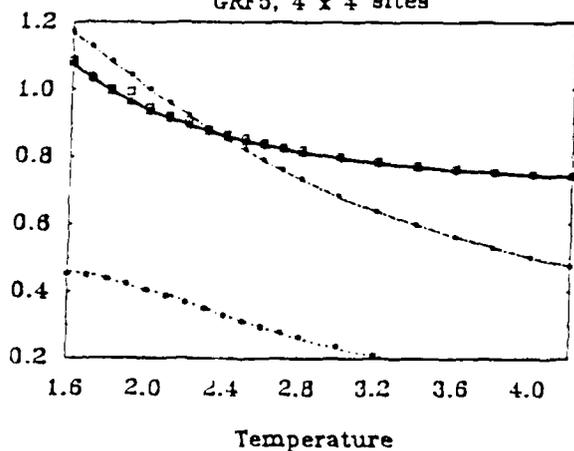


Figure 2E

GRF5, 4 x 4 sites



Figures 2A-2E: Comparison of the exact and estimated partition function, negative internal energy and specific heat for the five GRF models of Table II ( $K_1 = K_2 = 1600$ ).

$$\begin{aligned}
 \text{—} & \frac{1}{4^2} \ln Z_{4,4}, & + + & \frac{1}{4^2} \ln \hat{Z}_{4,4,P^{(4)}}(K), \\
 \square \square & \frac{1}{4^2} \ln \hat{Z}_{4,4,P^{\circ}}(K), & \Delta \Delta & \frac{1}{4^2} \ln \hat{Z}_{4,4,P^{\circ}}(K), \\
 \text{—} & -E_{4,4}(T), & \bullet \bullet & -\hat{E}_{4,4}(T; K_1, K_2), \\
 \text{.....} & C_{4,4}(T), & + + & \hat{C}_{4,4}(T; K_1, K_2).
 \end{aligned}$$

Figure 3A

GRF1, 4 x 4 sites

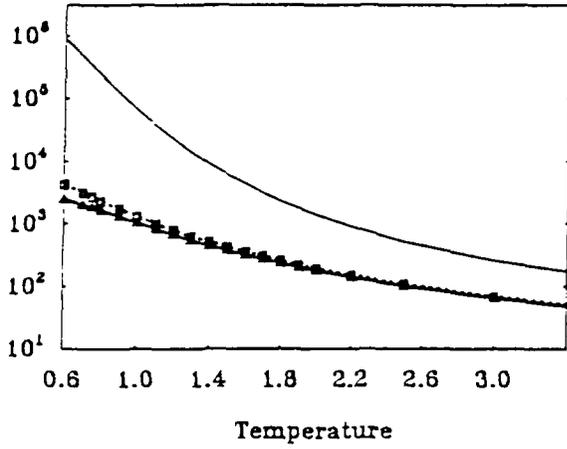


Figure 3B

GRF2, 4 x 4 sites

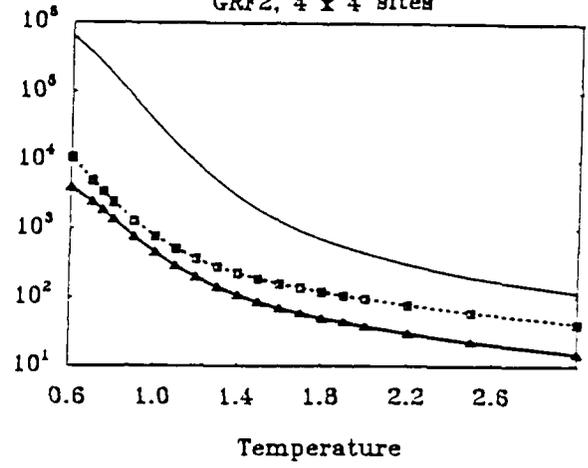


Figure 3C

GRF3, 4 x 4 sites

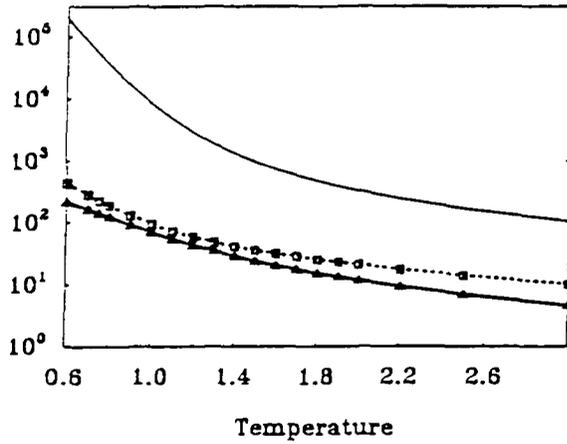


Figure 3D

GRF4, 4 x 4 sites

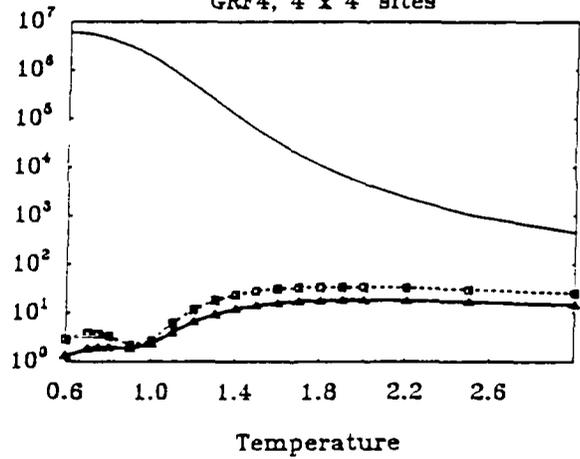
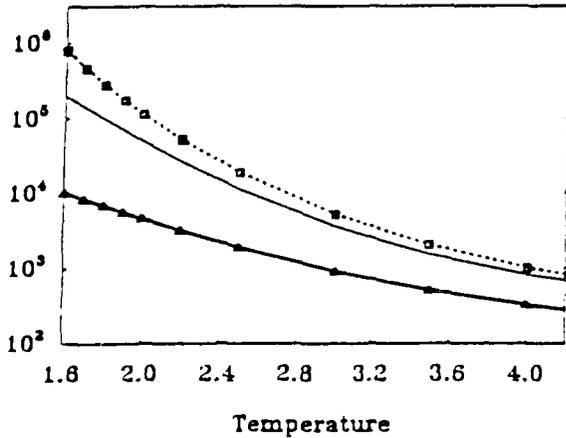


Figure 3E

GRF5, 4 x 4 sites



Figures 3A-3E: Number of iterations  $K_p$  of Algorithm I necessary to obtain

$Pr[\hat{Z}_{4,4,P}(K) \in (0.8Z_{4,4}, 1.2Z_{4,4})] \approx 0.95$ ,  
for  $K \geq K_p$ , for the five GRF models of Table II.

- $K_{p,nd}$  - i.i.d. choice,
- $K_{p,1}$  - Method 1,
- ▲-----▲  $K_{p,2}$  - Method 2.

Figure 4A

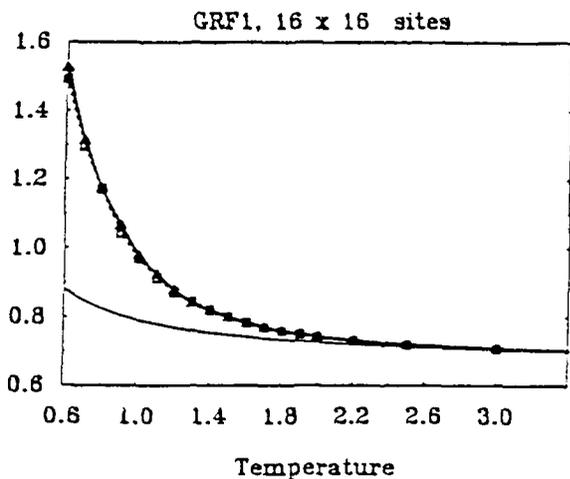


Figure 4B

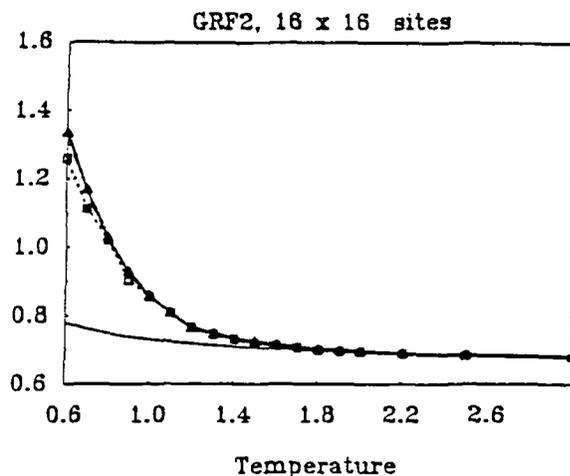


Figure 4C

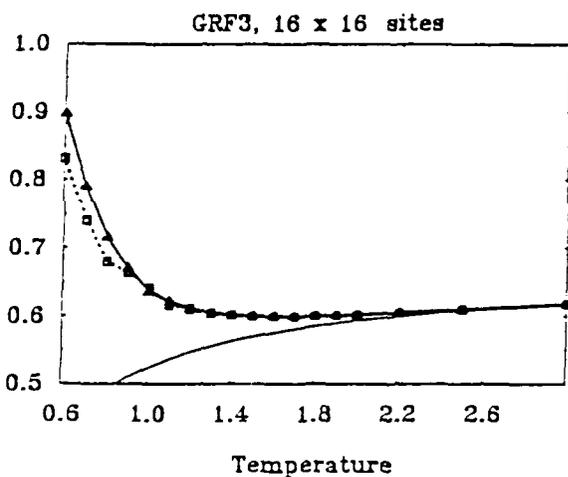


Figure 4D

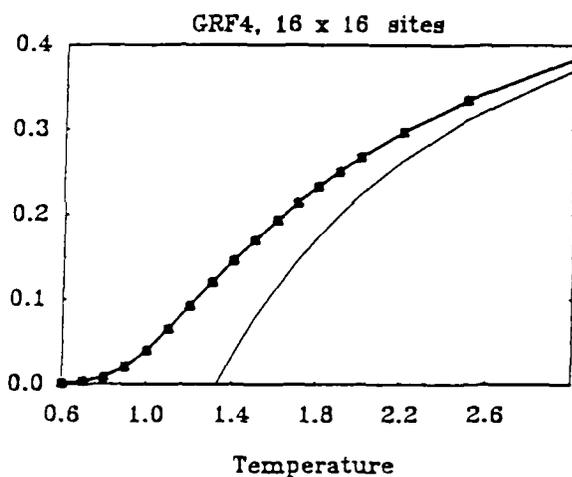
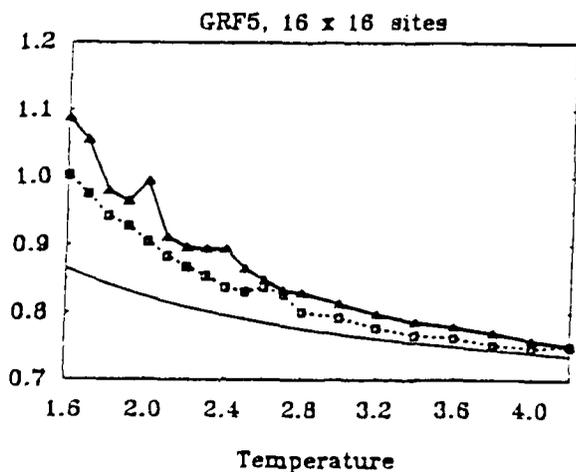


Figure 4E



Figures 4A-4E: Monte-Carlo estimation of the partition function of the five GRF models of Table II, using three different sampling schemes ( $K = 16^2 \times 10^4$ ).

- $\frac{1}{16^2} \ln \hat{Z}_{16,16,P^{iid}}(K)$  - *i.i.d. choice.*
- ...-□  $\frac{1}{16^2} \ln \hat{Z}_{16,16,P^*}(K)$  - *Method 1.*
- ▲-...-▲  $\frac{1}{16^2} \ln \hat{Z}_{16,16,P^*}(K)$  - *Method 2.*

Figure 5A

GRF1, 32 x 32 sites

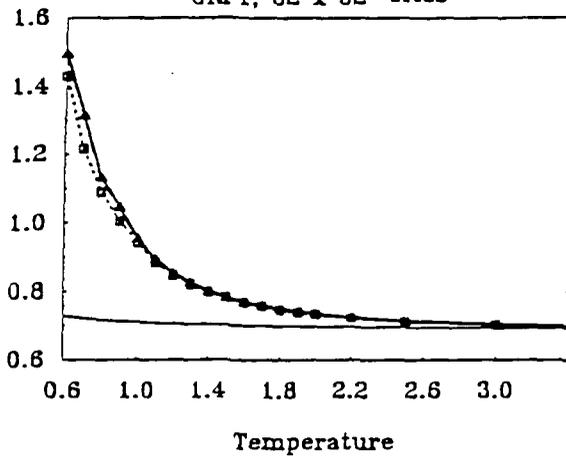


Figure 5B

GRF2, 32 x 32 sites

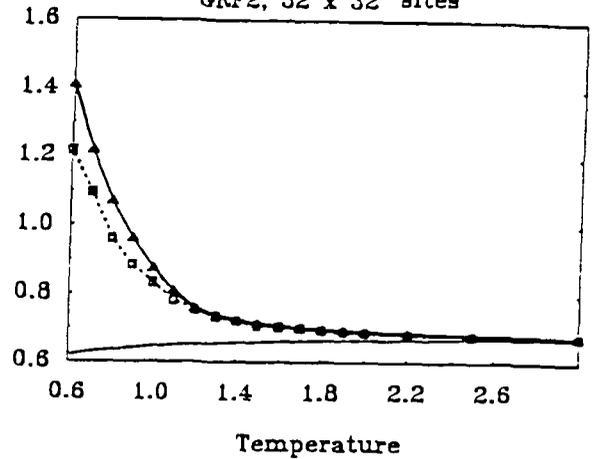


Figure 5C

GRF3, 32 x 32 sites

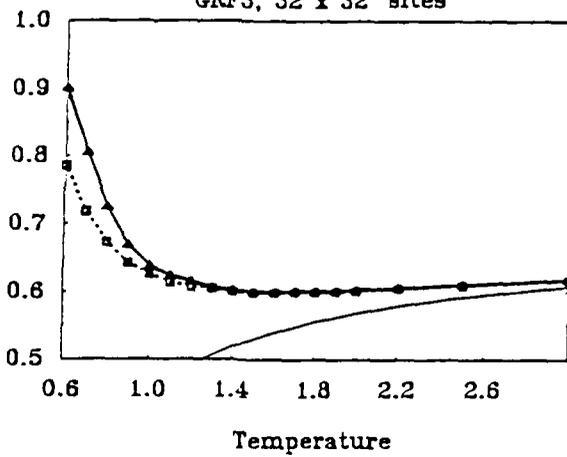


Figure 5D

GRF4, 32 x 32 sites

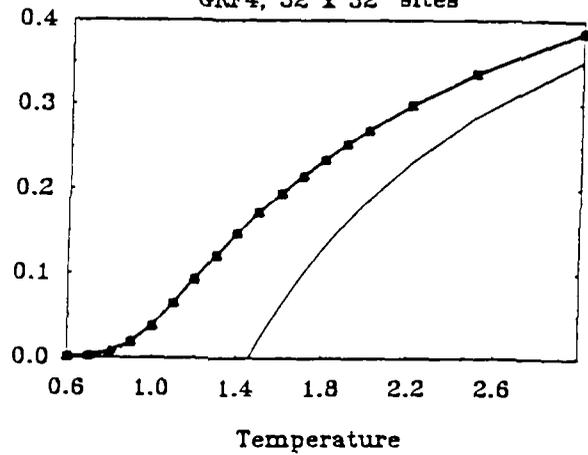
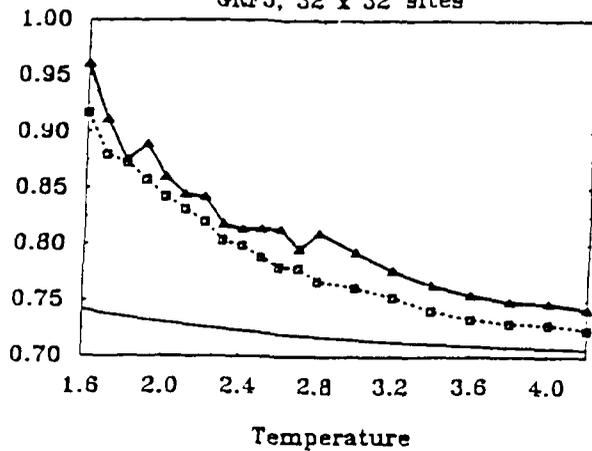


Figure 5E

GRF5, 32 x 32 sites



Figures 5A-5E: Monte-Carlo estimation of the partition function of the five GRF models of Table II, using three different sampling schemes ( $K = 32^2 \times 10^4$ ).

—  $\frac{1}{32^2} \ln \hat{Z}_{32,32,P^{i.i.d.}}(K)$  - *i.i.d. choice.*

□-...-□  $\frac{1}{32^2} \ln \hat{Z}_{32,32,P^*}(K)$  - *Method 1.*

▲-...-▲  $\frac{1}{32^2} \ln \hat{Z}_{32,32,P^-}(K)$  - *Method 2.*

Figure 6A

GRF1, 64 x 64 sites

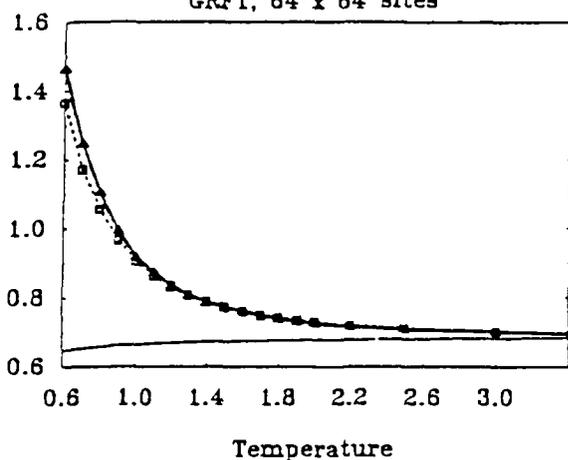


Figure 6B

GRF2, 64 x 64 sites

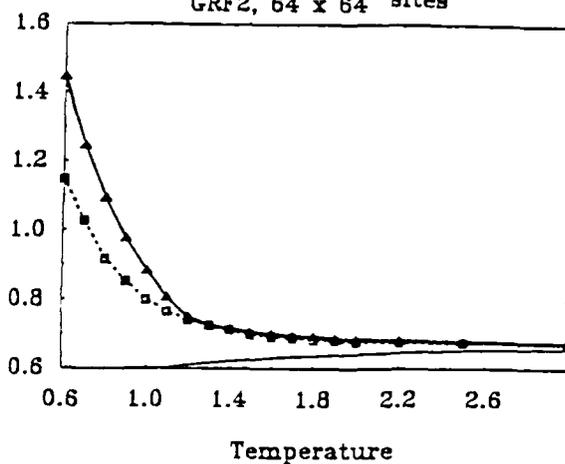


Figure 6C

GRF3, 64 x 64 sites

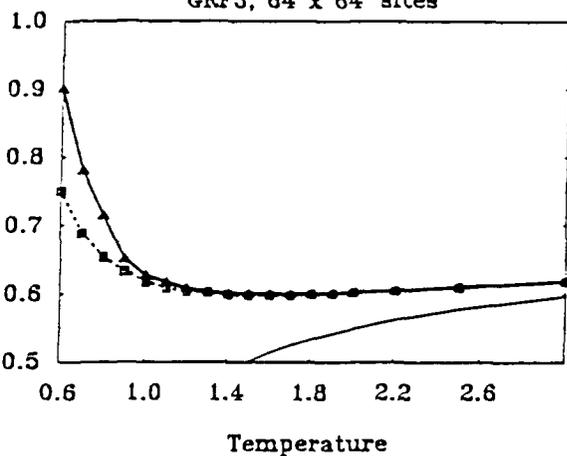


Figure 6D

GRF4, 64 x 64 sites

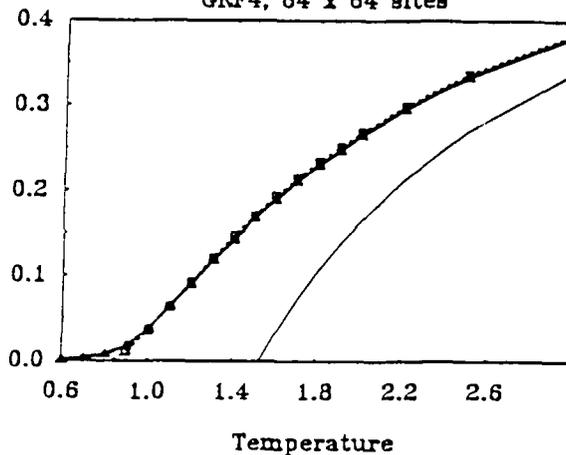
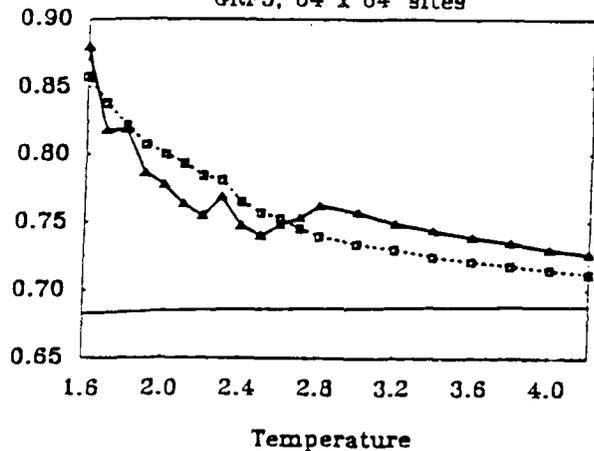


Figure 6E

GRF5, 64 x 64 sites



Figures 6A-6E: Monte-Carlo estimation of the partition function of the five GRF models of Table II, using three different sampling schemes ( $K = 64^2 \times 10^4$ ).

- $\frac{1}{64^2} \ln \hat{Z}_{64,64,P_{iid}}(K)$  - *i.i.d. choice.*
- ····· □  $\frac{1}{64^2} \ln \hat{Z}_{64,64,P^{\bullet}}(K)$  - *Method 1.*
- △ ····· △  $\frac{1}{64^2} \ln \hat{Z}_{64,64,P^{\circ}}(K)$  - *Method 2.*

Figure 7A

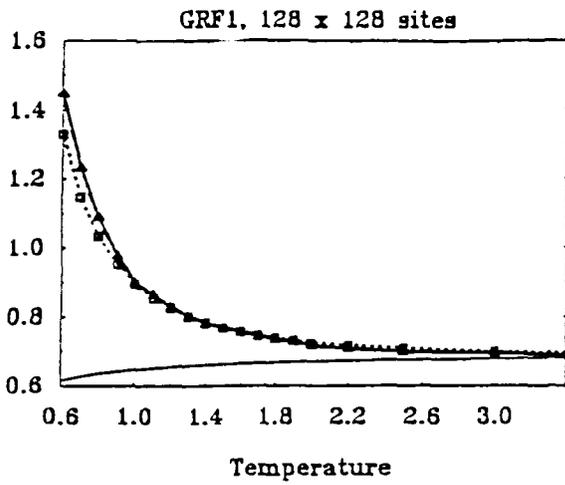


Figure 7B

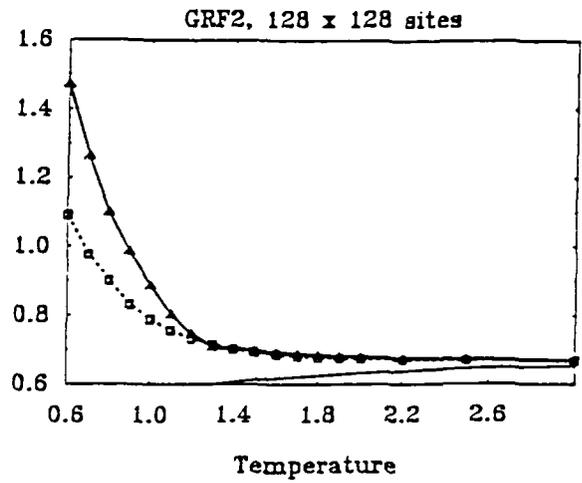


Figure 7C

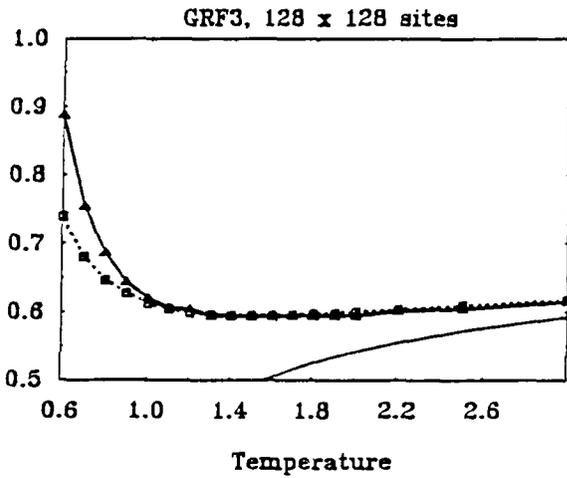


Figure 7D

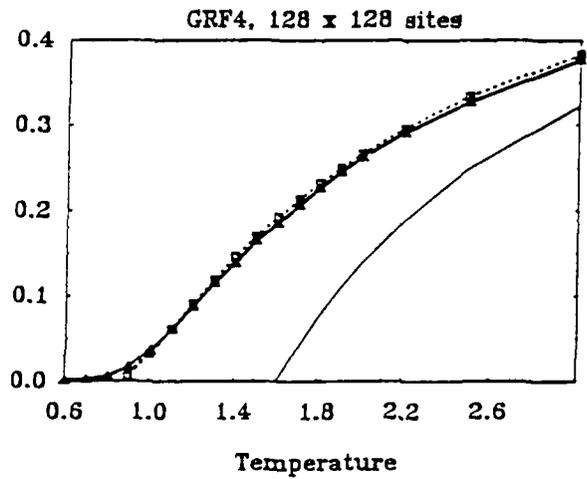
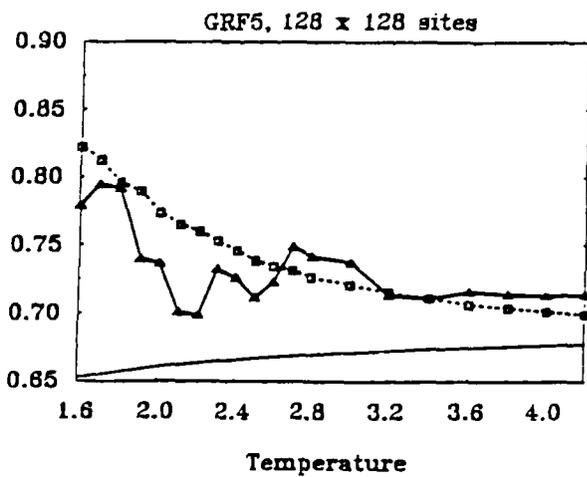


Figure 7E



Figures 7A-7E: Monte-Carlo estimation of the partition function of the five GRF models of Table II, using three different sampling schemes ( $K = 128^2 \times 10^4$ ).

- $\frac{1}{128^2} \ln \hat{Z}_{128,128,P^{i.i.d.}}(K)$  - *i.i.d. choice.*
- $\frac{1}{128^2} \ln \hat{Z}_{128,128,P^{\bullet}}(K)$  - *Method 1.*
- ▲---▲  $\frac{1}{128^2} \ln \hat{Z}_{128,128,P^{\circ}}(K)$  - *Method 2.*

Figure 8A

GRF1

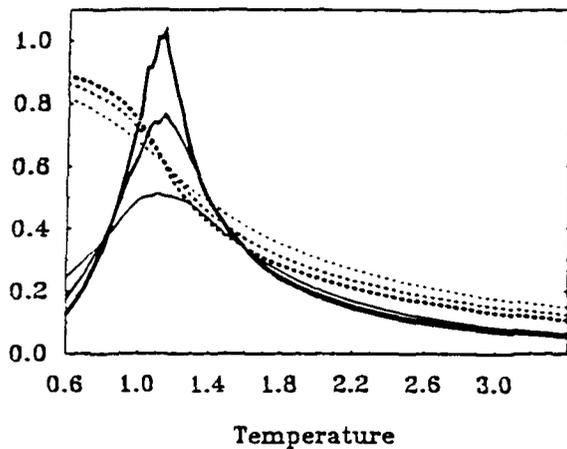


Figure 8B

GRF2

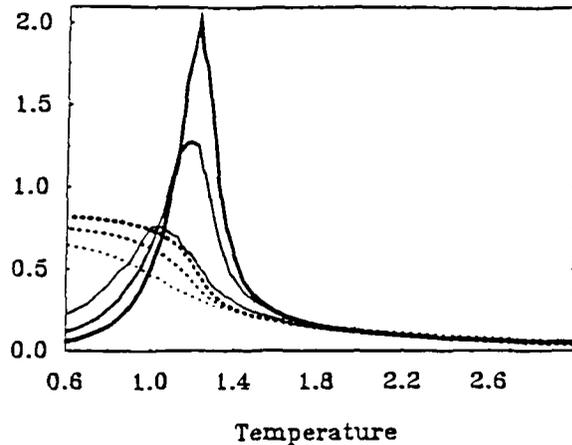


Figure 8C

GRF3

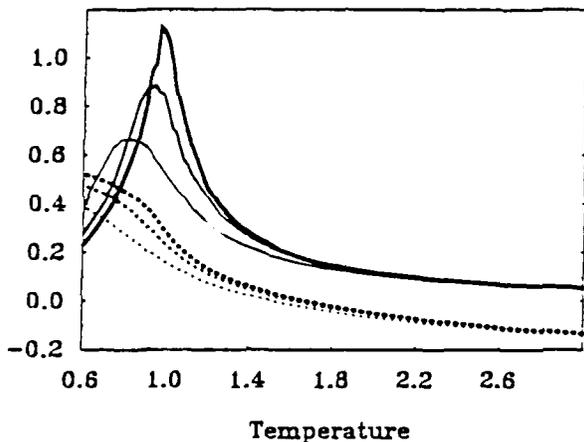


Figure 8D

GRF4

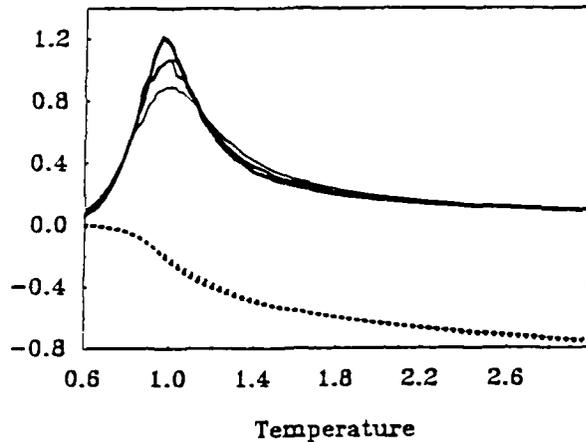
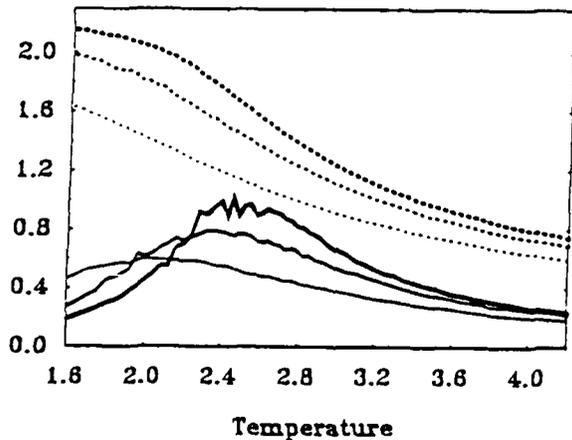


Figure 8E

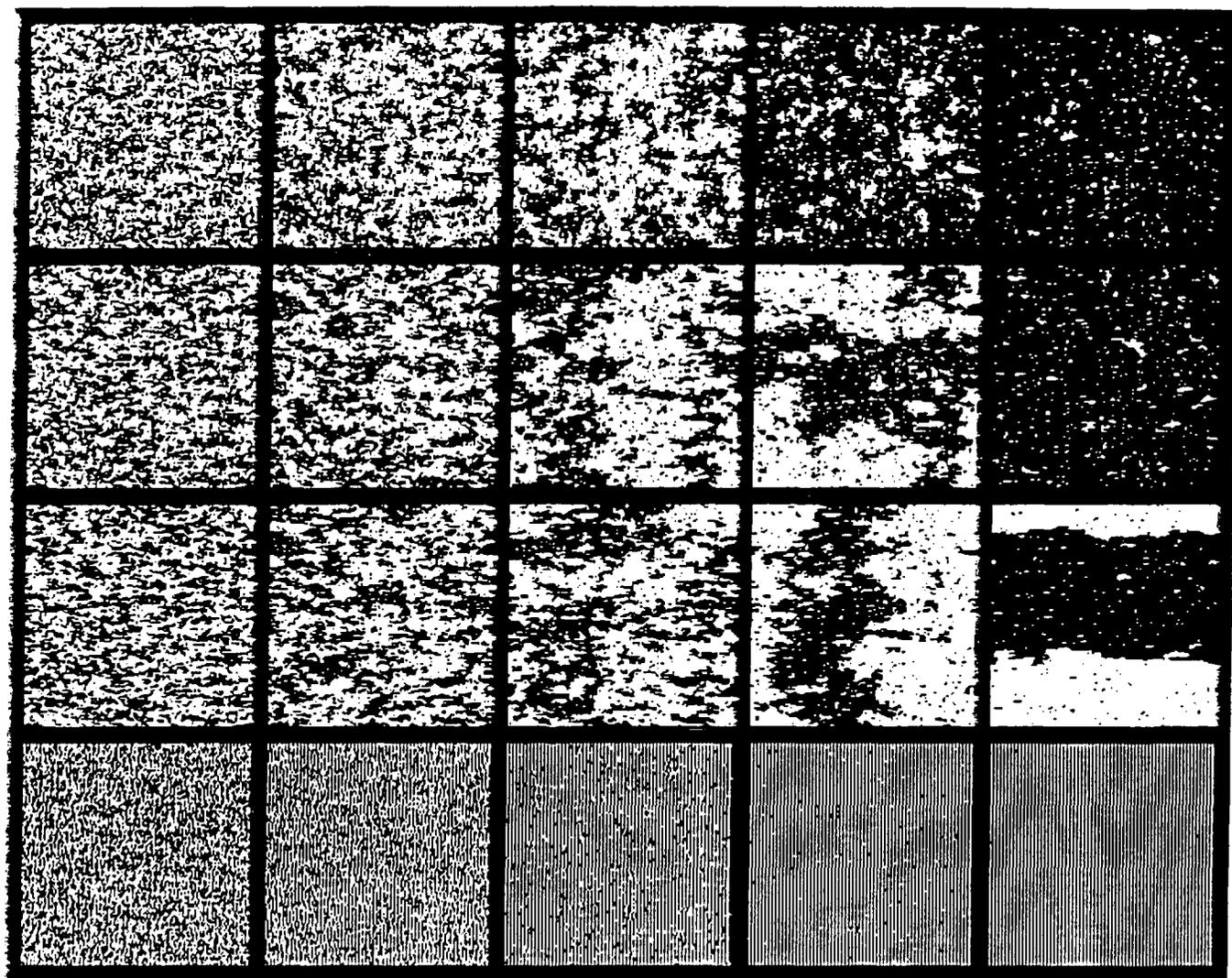
GRF5



Figures 8A-8E: Monte-Carlo estimates of

- .....  $-E_{8,8}(T)$ ,      —  $C_{8,8}(T)$ ,
- .....  $-E_{16,16}(T)$ ,      —  $C_{16,16}(T)$ ,
- .....  $-E_{32,32}(T)$ ,      —  $C_{32,32}(T)$ ,

for the five GRF models of Table II. In all cases  $K_1 = K_2 = 10,000 MN$ .

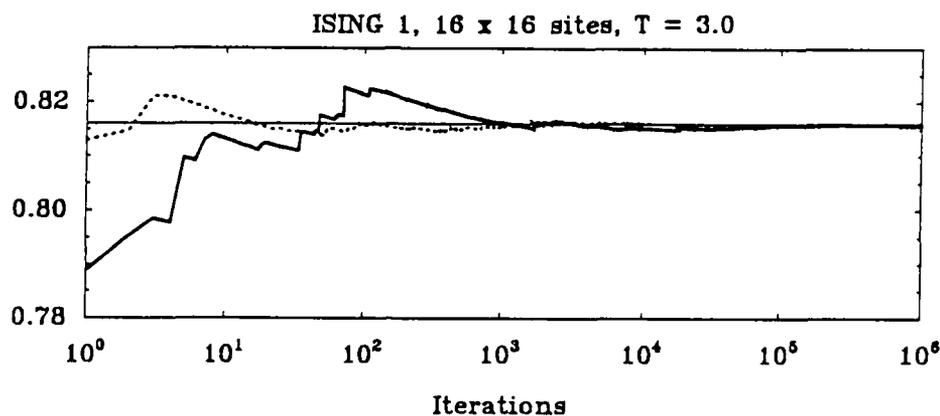


**Figure 9:** Realizations of  $128 \times 128$  site Ising models. Each row depicts an Ising model (from top to bottom: Ising 1, Ising 2, Ising 3, Ising 4) at five different temperatures (from left to right:  $T_0$ ,  $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ ). See Table III for details.

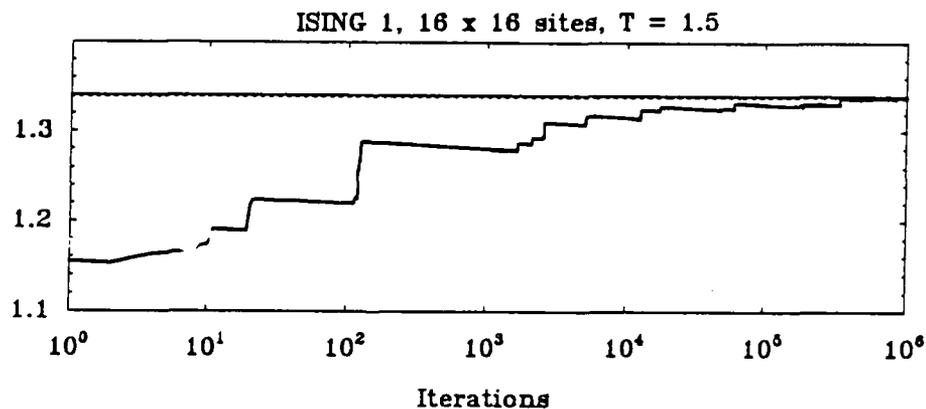
**Table III:** Interaction parameters  $(A, B)$  of the Ising models of Fig. 9 for different temperatures  $T_i$ .

Reference name	Interactions $(A, B)$	Crit. temp. $T_c$	Temperatures of Fig. 9				
			$T_0$	$T_1$	$T_2$	$T_3$	$T_4$
Ising 1	(1.0, 1.0)	2.26918	4.00	3.00	2.50	2.27	2.00
Ising 2	(2.0, 1.0)	3.28204	5.00	4.00	3.50	3.28	3.00
Ising 3	(3.0, 1.0)	4.15617	6.00	5.00	4.50	4.15	3.50
Ising 4	(-1.0, 1.0)	2.26918	4.00	2.50	2.00	1.50	1.00

**Figure 10A**



**Figure 10B**



**Figures 10A, 10B:** Convergence behavior of the Monte-Carlo algorithm for the estimation of the partition function of the  $16 \times 16$  site Ising 1 model at  $T = 3.0$  (Figure 10A,  $T > T_c$ ), or at  $T = 1.5$  (Figure 10B,  $T < T_c$ ).

—  $\frac{1}{16^2} \ln Z_{16,16}$  — Exact result.

—  $\frac{1}{16^2} \ln \hat{Z}_{16,16,P^*}(K)$  — Method 1 (plotted versus  $K$ ).

.....  $\frac{1}{16^2} \ln \hat{Z}_{16,16,P^{**}}(K)$  — Method 2 (plotted versus  $K$ ).

Figure 11A

ISING 1, 32 x 32 sites

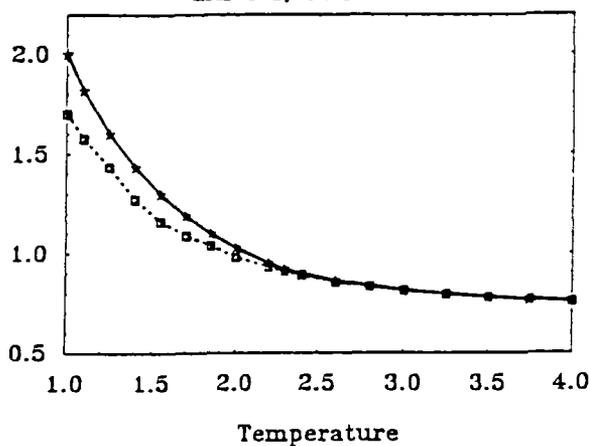


Figure 11B

ISING 2, 32 x 32 sites

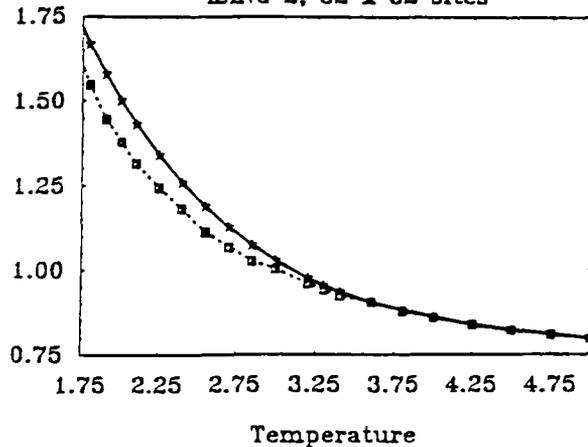


Figure 11C

ISING 3, 32 x 32 sites

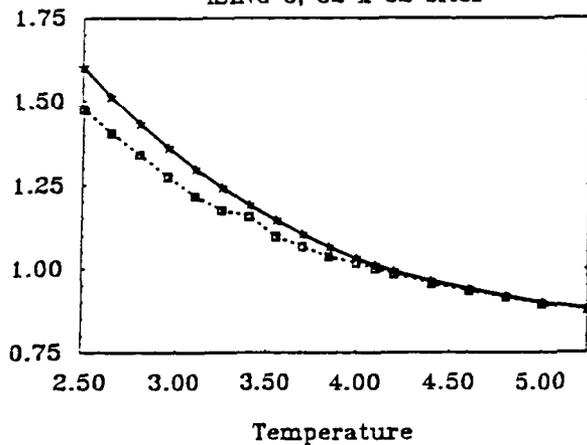
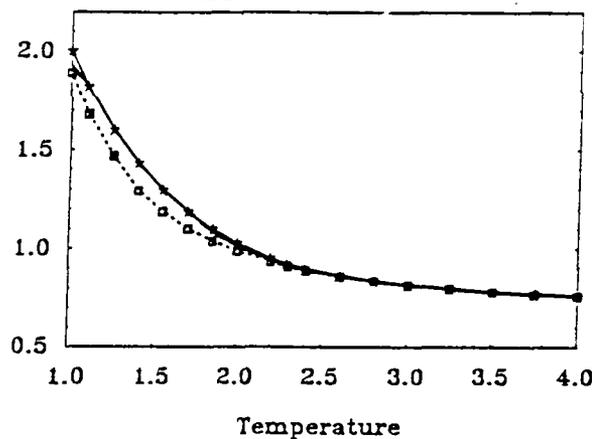


Figure 11D

ISING 4, 32 x 32 sites



Figures 11A-11D: Monte-Carlo estimation of the partition function of the four Ising models of Table III, considered on a 32x32 site lattice, by using two different sampling schemes ( $K = 32^2 \times 10^4$ ). Comparison with the exact partition function.

- $\times \text{---} \times$        $\frac{1}{32^2} \ln Z_{32,32}$       - Exact result.
- $\square \cdots \cdots \square$        $\frac{1}{32^2} \ln \hat{Z}_{32,32,P}(K)$       - Method 1.
- $\text{---}$        $\frac{1}{32^2} \ln \hat{Z}_{32,32,P^{\infty}}(K)$       - Method 2.

Figure 12A

ISING 1, 64 x 64 sites

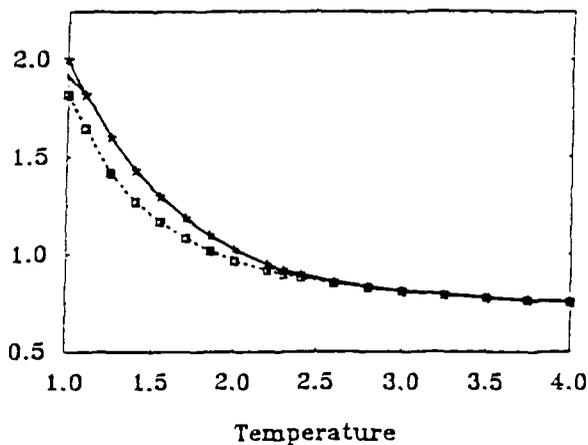


Figure 12B

ISING 2, 64 x 64 sites

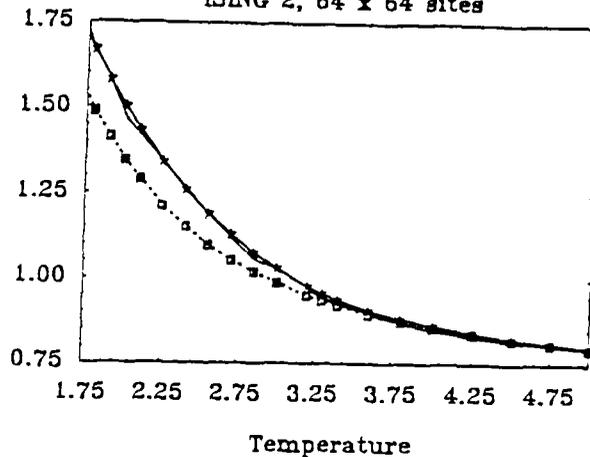


Figure 12C

ISING 3, 64 x 64 sites

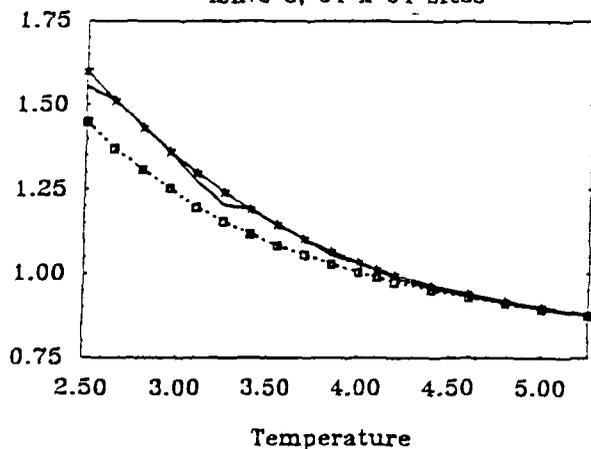
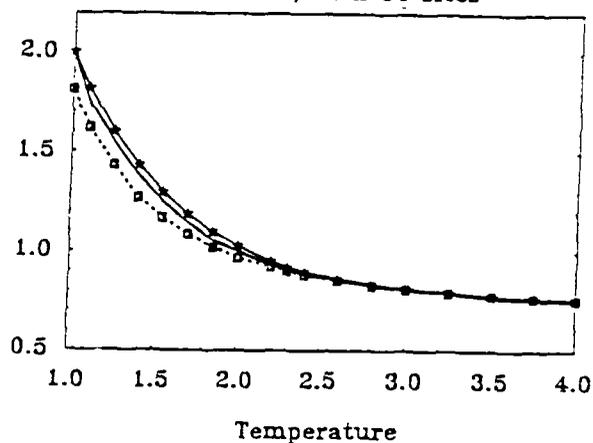


Figure 12D

ISING 4, 64 x 64 sites



Figures 12A-12D: Monte-Carlo estimation of the partition function of the four Ising models of Table III considered on a  $64 \times 64$  site lattice, by using two different sampling schemes ( $K = 64^2 \times 10^4$ ). Comparison with the exact partition function.

x ——— x	$\frac{1}{64^2} \ln Z_{64,64}$	- Exact result.
□ ····· □	$\frac{1}{64^2} \ln \hat{Z}_{64,64,P^{\bullet}}(K)$	- Method 1.
—————	$\frac{1}{64^2} \ln \hat{Z}_{64,64,P^{\circ}}(K)$	- Method 2.

Figure 13A

ISING 1, 128 x 128 sites

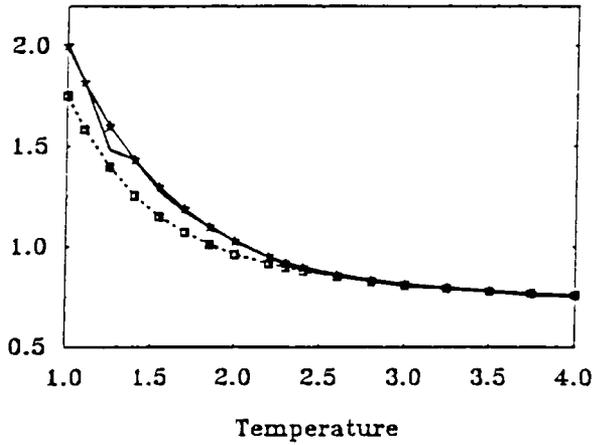


Figure 13B

ISING 2, 128 x 128 sites

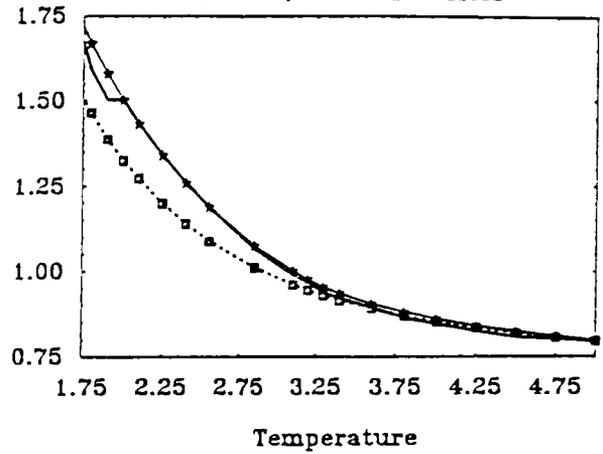


Figure 13C

ISING 3, 128 x 128 sites

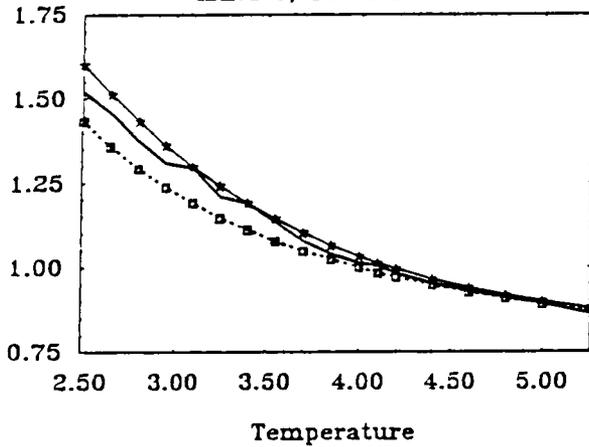
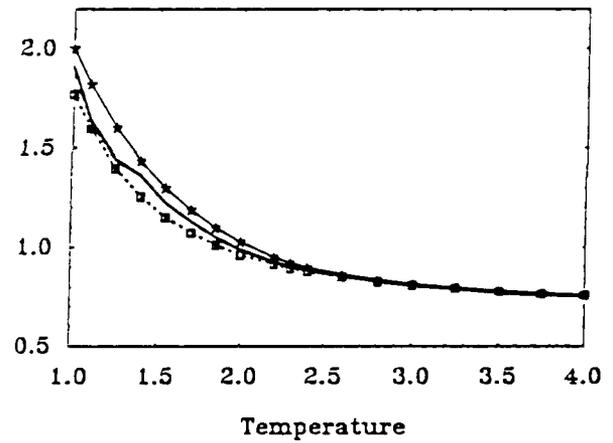


Figure 13D

ISING 4, 128 x 128 sites



Figures 13A-13D: Monte-Carlo estimation of the partition function of the four Ising models of Table III considered on a  $128 \times 128$  site lattice, by using two different sampling schemes ( $K = 128^2 \times 10^4$ ). Comparison with the exact partition function.

✕ ——— ✕	$\frac{1}{128^2} \ln Z_{128,128}$	- Exact result.
□ ····· □	$\frac{1}{128^2} \ln \hat{Z}_{128,128, P^{\bullet}}(K)$	- Method 1.
—————	$\frac{1}{128^2} \ln \hat{Z}_{128,128, P^{\circ}}(K)$	- Method 2.

Figure 14A

ISING 1, 32 x 32 sites

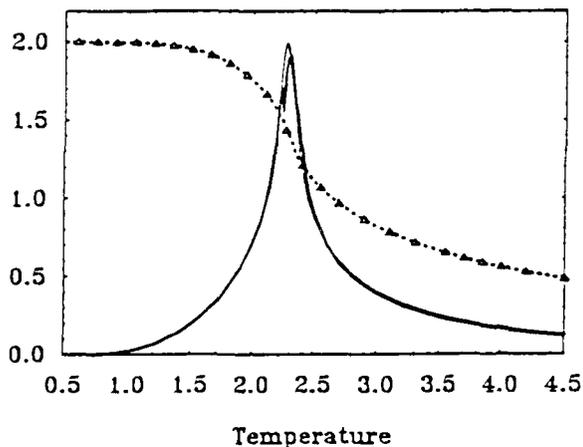


Figure 14B

ISING 2, 32 x 32 sites

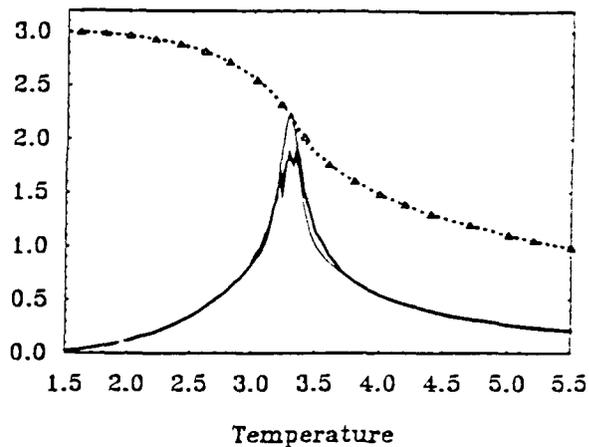


Figure 14C

ISING 3, 32 x 32 sites

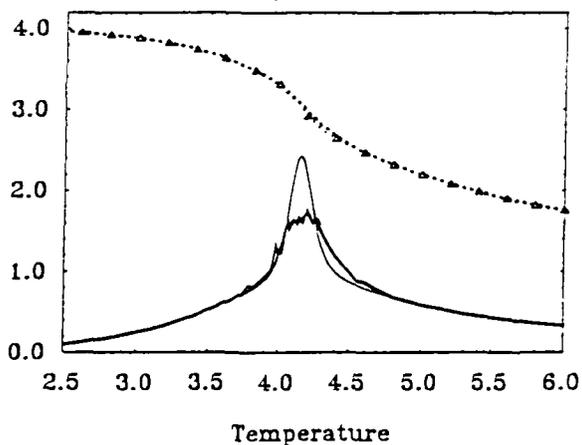
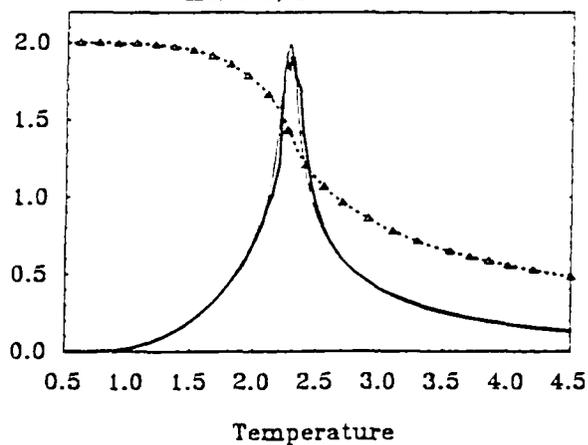


Figure 14D

ISING 4, 32 x 32 sites



Figures 14A-14D: Comparison of the exact and Monte-Carlo estimates of the internal energy and specific heat for the four Ising models of Table III, considered on a 32x32 site lattice ( $K_1 = K_2 = 10,000 \times 32^2$ ).

$\Delta \cdots \cdots \Delta$	$-E_{32,32}(T),$	$\cdots \cdots \cdots$	$-\hat{E}_{32,32}(T; K_1, K_2),$
$\text{—}$	$C_{32,32}(T),$	$\text{—}$	$\hat{C}_{32,32}(T; K_1, K_2).$

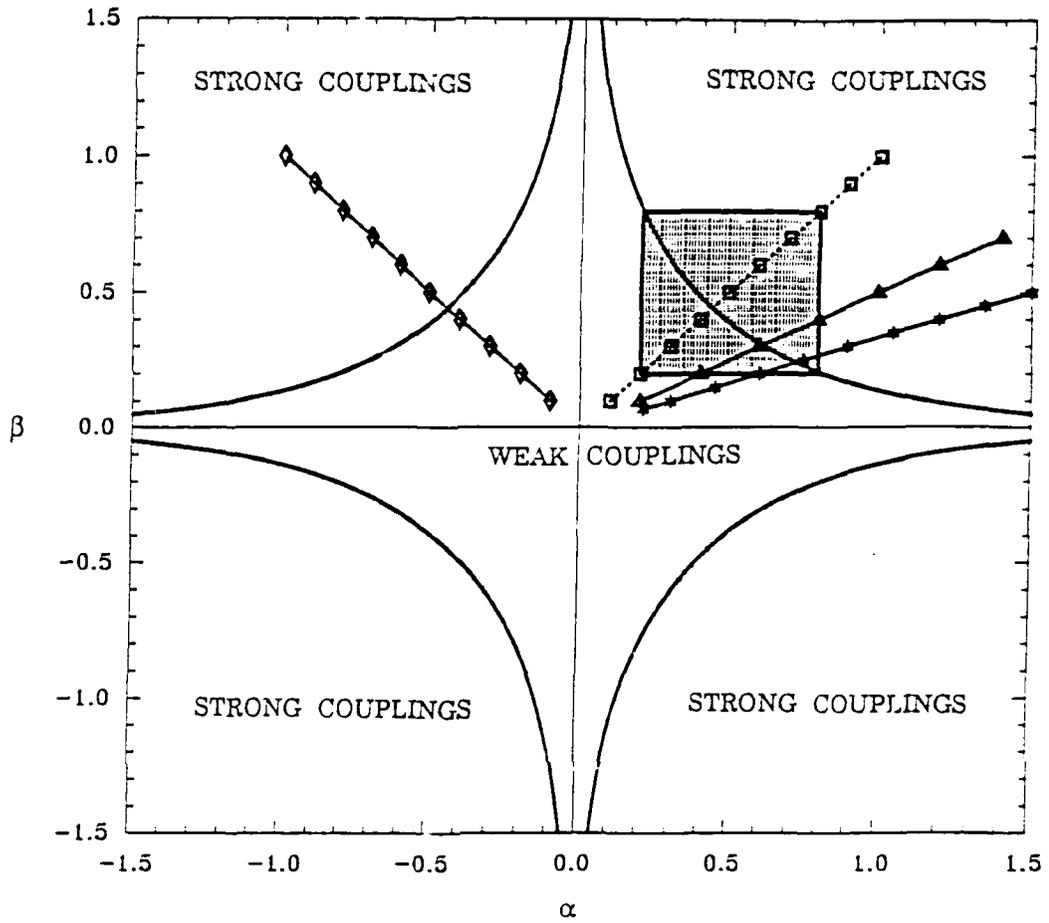


Figure 15: Areas of strong and weak couplings for the two-parameter Ising model. The regions which correspond to the parameter values considered in our simulation experiments are shown.

- 
Critical region: Separates areas of weak and strong couplings;
- 
Ising 1 model;
- 
Ising 2 model;
- 
Ising 3 model;
- 
Ising 4 model;
- 
the shaded area corresponds to the parameter values used for the results depicted in Figs. 16 and 17.

Figure 16A

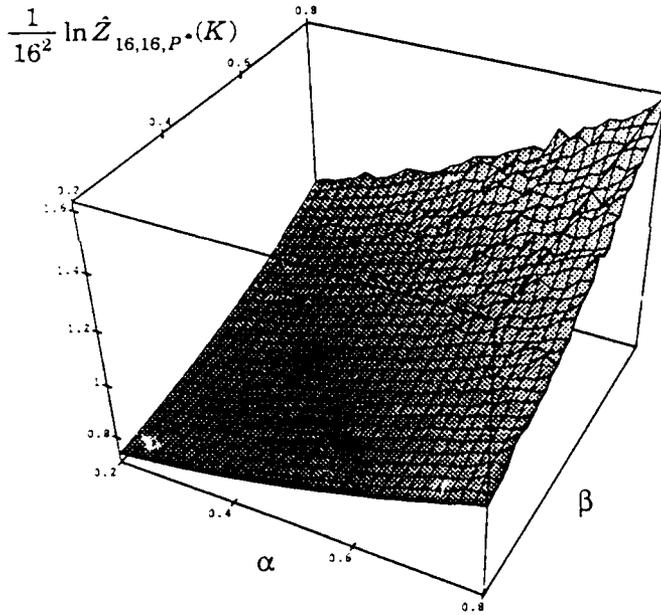
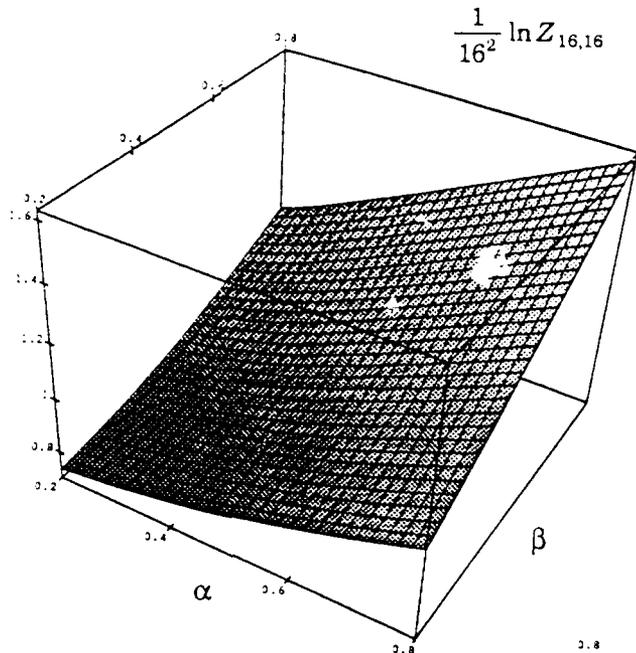


Figure 16B

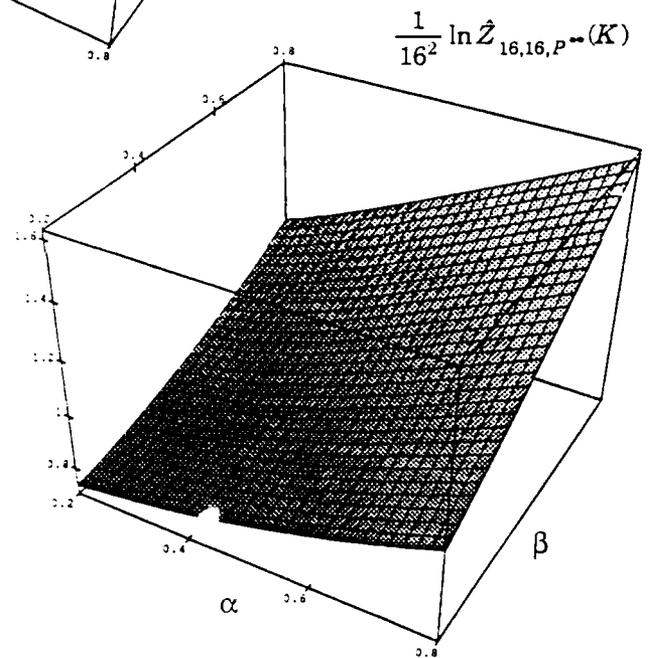


Figure 16C

Figures 16A-16C: Comparison of the exact and the Monte-Carlo estimated partition function of a  $16 \times 16$  site Ising model with parameters  $0.2 < \alpha < 0.8$  and  $0.2 < \beta < 0.8$  ( $K = 16^2 \times 10^4$ ).

Figure 17A

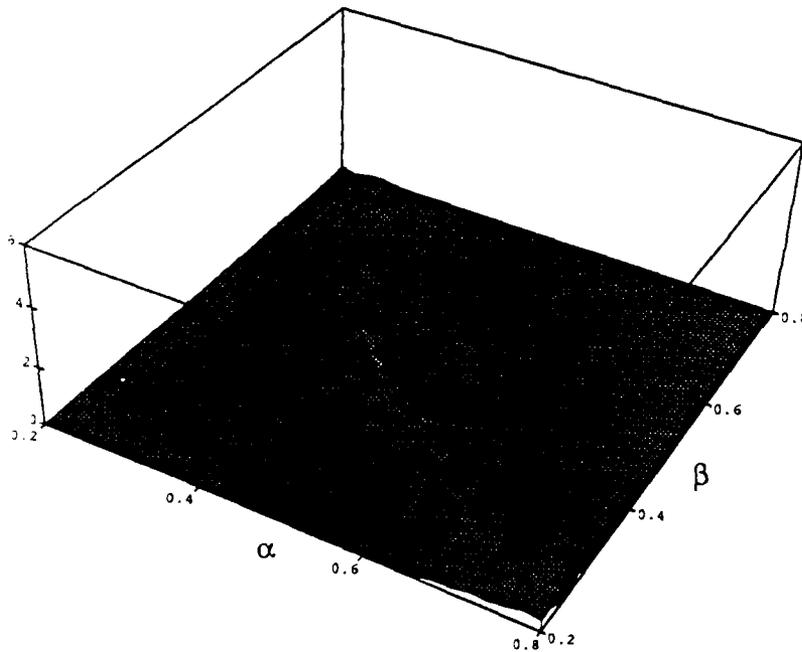
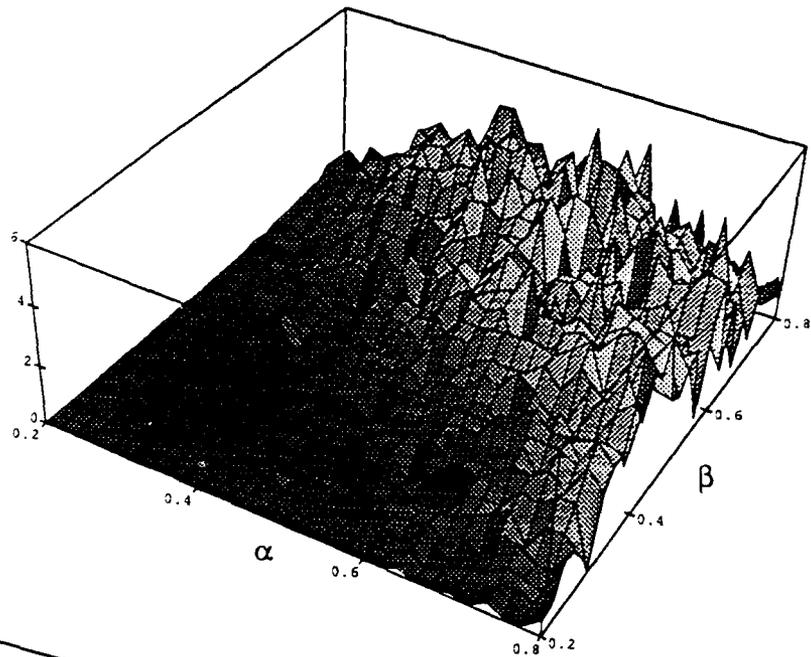


Figure 17B

**Figures 17A, 17B:** Error (%) of the Monte-Carlo estimate of the partition function for the Ising model considered in Figs. 16A-C.

$$\text{Figure 17A: } 100 \frac{|\ln \hat{Z}_{16,16,P^*(K)} - \ln Z_{16,16}|}{\ln Z_{16,16}} \quad - \text{Method 1.}$$

$$\text{Figure 17B: } 100 \frac{|\ln \hat{Z}_{16,16,P^{**}(K)} - \ln Z_{16,16}|}{\ln Z_{16,16}} \quad - \text{Method 2.}$$

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>			1b. RESTRICTIVE MARKINGS			
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT			
7d. DECLASSIFICATION/DOWNGRADING SCHEDULE			<b>UNLIMITED</b>			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)			
6a. NAME OF PERFORMING ORGANIZATION  The Johns Hopkins University		6b. OFFICE SYMBOL <i>(If applicable)</i>		7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State and ZIP Code) Department of Electrical and Computer Engr. Image Analysis and Communications Laboratory Baltimore, MD 21218			7b. ADDRESS (City, State and ZIP Code)			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Department of the Navy		8b. OFFICE SYMBOL <i>(If applicable)</i>		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State and ZIP Code) Office of the Chief of Naval Research 800 North Quincy Street, Code 1511:CLC Arlington, Virginia 22217-5000			10. SOURCE OF FUNDING NOS. <b>N00014-90-J-1345</b>			
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
11. TITLE (Include Security Classification) <b>Stochastic Simulation Techniques for Partition Function Approximation of Gibbs Random Filed Images)</b>						
12. PERSONAL AUTHOR(S) G. Potamianos and J. Goutsias						
13a. TYPE OF REPORT Technical Report		13b. TIME COVERED FROM <b>01/01/90</b> TO <b>06/30/91</b>		14. DATE OF REPORT (Yr., Mo., Day) June, 1991		15. PAGE COUNT 68
16. SUPPLEMENTARY NOTATION						
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)			
FIELD	GROUP	SUB. GR.	Gibbs Random Fields, Monte Carlo Simulations, Partition Function, Image Processing			
19. ABSTRACT (Continue on reverse if necessary and identify by block number)						
<p>A Monte-Carlo simulation technique for the calculation of the partition function of a general Gibbs random field is presented. We show that the partition function of a general Gibbs random field is equivalent to an expectation. This observation allows us to develop an importance sampling approach for estimating this expectation by using Monte-Carlo simulations. Two different methods are proposed for this task. We show that the estimators are unbiased and consistent. Computations are performed iteratively, by using a simple, Metropolis-like, Monte-Carlo algorithm with remarkable success, as it is demonstrated by our simulations. Our work concentrates on binary, second-order Gibbs random fields defined on a rectangular lattice. However, the proposed methods can be easily extended to more general Gibbs random fields and, therefore, can become quite useful in many scientific areas, such as biology, statistical mechanics and image processing. Furthermore, a potential contribution of our technique to optimally estimating the parameters of a general Gibbs random field from a given realization via a maximum-likelihood approach is anticipated.</p>						
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>				21. ABSTRACT SECURITY CLASSIFICATION		
22a. NAME OF RESPONSIBLE INDIVIDUAL John Goutsias			22b. TELEPHONE NUMBER <i>(Include Area Code)</i> (301) 338-7871		22c. OFFICE SYMBOL	