The Annals of Applied Probability
1993, Vol. 3, No. 1, 137–153

CONVERGENCE OF SOME PARTIALLY PARALLEL GIBBS SAMPLERS WITH ANNEALING[1]

BY PABLO A. FERRARI, ARNOLDO FRIGESSI AND ROBERTO H. SCHONMANN

Universidade de São Paulo, Istituto per le Applicazioni del Calcolo del CNR and University of California, Los Angeles

In this paper we consider the Gibbs sampler dynamics with simulated annealing and partially parallel updating scheme, as proposed by Trouvé. It is known that in some cases the support of the limiting measure does not coincide with the set of global maxima of the underlying energy function. We provide some new simple examples of this undesirable behavior. However, we also prove that for one-dimensional binary models with nearest neighbor interaction the algorithm works "generically." We prove also that for two-dimensional models with nearest neighbor ferromagnetic constant interactions the algorithm works.

1. Introduction. Simulated annealing is a very well known algorithm for global optimization. Its convergence properties have been intensively studied, both from a theoretical point of view and for practical purposes (see, e.g., Aarts and van Laarhoven [1] and Hwang and Sheu [9] and references therein). It is all-purpose, and this is its main advantage; it has been applied to many hard problems which could not be solved in a satisfactory way otherwise. The most serious drawback lies in its very slow convergence rate. Simulated annealing is a stochastic iterative algorithm that converges (in a proper sense, to be made precise below) to the set of points which are the global maxima of a given function. But this convergence occurs if and only if a certain parameter, called temperature, is decreased slowly enough. This cooling schedule is so slow, that in most practical cases it has to be replaced by a faster one, although convergence to the extrema cannot be proven anymore. The parallelization of simulated annealing is therefore of growing interest, as parallel architectures become more and more available.

The simulated annealing algorithm that we consider in this paper is based on a discrete time Markov chain, whose state space coincides with the domain of the energy function H to be maximized. The transition matrix is defined so that the chain converges weakly to the uniform distribution over the maxima of H. Typical problems to which it is applied arise, for instance, in statistical mechanics (see Sokal [11]), neural networks and learning (Azencott [2]), image restoration (Geman [7]) and combinatorial optimization (Aarts and

Received May 1991; revised February 1992.

[1]Research partially supported by Fundação de Apoio à Pesquisa do Estado de São Paulo, Consiglio Nazionale delle Ricerche, N.A.T.O Advanced Fellowship 215.24.01, CNPq, Projeto BID-USP and NSF Grant DMS-91-00725.

AMS 1991 *subject classifications*. Primary 60J10, 60K35; secondary 65C05, 68Q10.

Key words and phrases. Simulated annealing, parallel algorithms, support of the limiting measure, global optimization, spin glasses, voter model.

van Laarhoven [1]). The common feature of these problems is a very large but finite state space, which cannot be enumerated in practice. We are concerned with lattice based problems, because of their simple parallelization. By this we mean that H is defined on $\Omega = L^{\Lambda}$, where Λ is a finite sublattice of $\mathbb{Z}^d$ and L is a finite set of possible values (colors) taken at each site of Λ. The algorithms we consider act locally, by updating single sites, according to a specified local transition rule. In this paper we concentrate on the Gibbs sampler proposed by Geman and Geman in [6]. The transition rule is given by the conditional probability distribution of the site being updated, given all other variables. The algorithm can be asynchronous, when sites are updated one at a time, for example, according to a random sweep strategy along the lattice; or the dynamics can be synchronous, when all sites are updated simultaneously. In the asynchronous case, when the temperature goes to zero slowly enough, the process converges to a convex combination of the measures that concentrate mass on the states that maximize H. Unfortunately, in the synchronous case, the algorithm does not always converge to that measure (see Azencott [2] and Dawson [3]) and is therefore not always useful.

Trouvé [13] proposes a very interesting algorithm, that lies between the synchronous and the asynchronous ones: At each time, each site is updated independently with a probability τ, while it does not change its current value with probability $1 - \tau$. For $\tau = 1$ this is the synchronous algorithm, and for τ close to 0 it may be considered to be similar to the asynchronous one. In terms of parallelization, this partially parallel algorithm is more profitable for values of τ close to one. Trouvé proves, quite unexpectedly, that the support of the limit distribution of the partially parallel Gibbs sampler with annealing does not depend on the value of $\tau \in (0, 1)$. Hence the crucial question arises: Is this support the set Ω^* of global maxima of H, or is it another set, as in the case of the fully parallel dynamics? The answer is not complete yet, but some information is available. On one hand, there are examples for which the limiting support is not Ω^* for $\tau > 0$: see, for instance, Trouvé [12] or Section 2 of this paper. On the other hand, numerical experiments done by Gaudron and Trouvé [5] with two dimensional spin glasses indicate good agreement between the limiting support of H, as $t \to \infty$, when sequential or partially parallel updating schemes are used. These numerical results allow one to be quite optimistic concerning the practical application of the partially parallel scheme.

This paper aims at clarifying the apparent contradiction pointed out above. We will show that in fact the counterexamples are not generic, but instead depend on a careful choice of the interactions (i.e., the constant parameters that appear in the definition of the energy function). We will see that if for the same lattice the interactions are chosen at random from continuous distributions, then almost surely the resulting model is well behaved, in the sense that the corresponding partially parallel Gibbs sampler, with logarithmic annealing schedule, has a limiting distribution concentrated on Ω^*. The models which we treat in this paper are very simple and should be considered as toy models if compared to those really interesting for applications. However our investigation points out to some important properties of the interactions which may cause the partially parallel algorithm to work or fail.

In the next section we introduce some notation and give the precise definitions; also some results obtained by Trouvé in [12] and [13] are recalled. Section 3 is devoted to the one-dimensional spin glass model with random interactions and two possible colors. Our results in this section are based on techniques borrowed from the theory of interacting particle systems (See Durrett [4], Griffeath [8] and Liggett [10]). In Section 4 we study the limiting support for the two-dimensional ferromagnetic Ising model. The analysis is based on estimates of hitting times for the underlying Markov chain. The analysis of the extrema for this model has no practical purpose, since they are known, but the type of arguments used in the proof may turn out to be relevant in other, more interesting, cases.

We end this introduction with a more practical remark: If the Gibbs sampler is implemented on real parallel machines, where a CPU is dedicated to each site, the random parallelization scheme produced by the Bernoulli site activation process could probably be avoided. In fact the CPUs will be only partially synchronized; hence a random parallelization would be "naturally" produced, in a pseudo-random fashion.

2. The partially parallel Gibbs sampler. We will construct Markov chains $(\eta_t)_{t=0,1,\dots}$ which take values on $\Omega := L^\Lambda$, where $\Lambda = \{1,\dots,N\}^d \subset \mathbb{Z}^d$, $d \geq 1$ and $L = \{+1, -1\}$. Consider the nearest neighbor energy function H on Ω given by

$$H(\eta) := \sum_{\langle x,y\rangle} J_{x,y}\eta(x)\eta(y), \tag{1}$$

where the sum is over neighboring pairs of sites in Λ. Free or periodic boundary conditions are assumed. The interactions $J_{x,y}$ may be chosen randomly, for example, as i.i.d. Gaussian random variables; however, after being chosen, the interactions are kept fixed. The Gibbs measure associated with the energy function $H(\cdot)$ at inverse temperature β is

$$\mu_\beta(\eta) := \frac{\exp(\beta H(\eta))}{Z},$$

where $Z = \sum_\eta \exp(\beta H(\eta))$.

Let $\Omega^* = \{\eta \in \Omega: H(\eta) \geq H(\xi), \forall\, \xi \in \Omega\}$.

In order to describe the Gibbs sampler dynamics, we define the local characteristics at site x by

$$q_x(\eta,\xi) = \mu_\beta(\sigma(x) = \xi(x) | \sigma(y) = \eta(y), y \neq x).$$

Next we introduce the different parallelization schemes.

Sequential dynamics. At each step choose a site uniformly at random and update it. The probability transition matrix is, for $\eta \neq \xi$,

$$Q_\beta(\eta,\xi) = \begin{cases} (1/|\Lambda|) q_x(\eta,\xi), & \text{if } \xi \in N(\eta,x) = \{\sigma\colon \sigma(y) = \eta(y), y \neq x\} \\ & \qquad \text{for some } x \in \Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

These dynamics are reversible with respect to the measure μ_β and hence the process, being irreducible and aperiodic, converges weakly to it.

Parallel dynamics. All sites are modified simultaneously. The probability transition matrix is

$$Q_{1,\beta}(\eta,\xi) = \prod_x q_x(\eta,\xi).$$

This process is typically not even irreducible, and μ_β is not one of its invariant measures; see, for example, Azencott [2]. To understand this important point, consider the case in which free boundary conditions are used: With parallel updating, the configurations on the even and odd sublattices (i.e., the sublattices on which the sum of the values of the coordinates are, respectively, even or odd) are independent at any given time if they were so at time zero.

Partially parallel dynamics [13]. In this case each site will be updated with probability τ, independently. The transition probability matrix is

$$Q_{\tau,\beta}(\eta,\xi) = \prod_x \left(\tau q_x(\eta,\xi) + (1-\tau)\mathbf{1}_{\{\eta(x)=\xi(x)\}}(\eta,\xi)\right). \tag{2}$$

For $\tau \in (0,1)$ the Markov chain with transition probabilities (2) is ergodic and converges to an invariant distribution $\mu_{\tau,\beta}$. Of course, for $\tau = 1$, $Q_{\tau,\beta}(\cdot,\cdot)$ gives rise to the parallel dynamics and for $\tau = 0$ it corresponds to the trivial dynamics in which the initial configuration never changes. However for small positive τ there are occasional flips of a single spin, and much less frequent flips of more than one spin at a time. It is easy to see that in the limit in which $\tau \to 0$, after rescaling time, the partially parallel dynamics approaches a continuous time Markov process whose embedded Markov chain is the sequential dynamics.

The limit

$$\lim_{\beta\to\infty} \mu_{\tau,\beta}$$

exists, as shown by Trouvé [13], and we will denote it by $\mu_{\tau,\infty}$.

Let $Q_{\tau,\infty} = \lim_{\beta\to\infty} Q_{\tau,\beta}$, which is also well defined as observed in [13], and consider the convex set of invariant measures

$$I_{\tau,\infty} = \{\mu\colon \mu Q_{\tau,\infty} = \mu\}.$$

Consider the invariance condition $\mu_{\tau,\beta} Q_{\tau,\beta} = \mu_{\tau,\beta}$, and take the limit in β, to see that

$$\mu_{\tau,\infty} \in I_{\tau,\infty}. \tag{3}$$

(See Proposition 2.14 of Chapter I in Liggett [10], where the fact above is shown to be true even for infinite systems and in continuous time.) Trouvé proves also in [13] that the *limiting support*

$$\mathrm{supp}(\mu_{\tau,\infty})$$

is independent of $\tau \in (0,1)$.

Take now a sequence of inverse temperatures $\{\beta_t\}_{t\in\mathbb{N}}$ increasing to infinity, and let η_t be the time-inhomogeneous Markov process with transition probabilities

$$P(\eta_{t+1} = \xi | \eta_t = \zeta) = Q_{\tau,\beta_t}(\zeta, \xi).$$

Recall that this process depends on τ, which is held fixed. The standard simulated annealing convergence theorem is proved for the partially parallel dynamics by Trouvé in [13]:

THEOREM 1. *Let $0 < \tau < 1$. There is a constant C (that depends on the energy H, τ and the number of sites) such that if β_t increases to infinity but $\beta_t \le C \ln t$, then*

$$\lim_{t\to\infty} \| P(\eta_t = * | \eta_0) - \mu_{\tau,\infty}(*) \| = 0,$$

where $\|\mu - \nu\| := \sum_\eta |\mu(\eta) - \nu(\eta)|$.

It is not difficult to construct examples for which the limiting support does not coincide with the set Ω^* of global maxima of $H(\cdot)$. Trouvé describes in [13] such an example, but his function $H(\cdot)$ does not have the form (1) and is rather artificial. The following is a simple example in which $H(\cdot)$ is of the form (1). Let $d = 1$, $N = 3$, consider periodic boundary conditions and let $J_{x,y} = -1$ for every pair of (neighboring) sites x and y. In other words, we are considering three spins coupled to each other with the same antiferromagnetic coupling. The set Ω^* contains the six configurations with two spins $+$ and one -1 and vice versa. On the other hand, the limiting support is the full set of all eight possible configurations, including the two configurations with all spins aligned. One can easily check this claim, by observing that enough entries of the transition matrix $Q_{\tau,\beta}$ remain positive in the limit as $\beta \to \infty$, so that $Q_{\tau,\infty}$ is irreducible. Hence the set $I_{\tau,\infty}$ has a unique element, which gives positive mass to all the configurations. The claim follows then from (3). Note that this argument relies on the fact that all the couplings $J_{x,y}$ have the same value. The mechanism which provides the nonvanishing probability of jumping from a configuration with exactly two alike spins to a configuration with three identical spins is the simultaneous update of the two parallel spins but not of the other one. Due to the fact that the three couplings are identical, this transition has probability $\tau^2(1-\tau)(1/2)^2$ for every value of β.

Now let us modify slightly the model so that the three interactions are still negative but with one of them being smaller than the others in absolute value. The configuration in which the spins connected by the smallest interaction are identical while the other one is different is the only one which maximizes the energy. However, in this case the probability of escaping from it vanishes as $\beta \to \infty$. In other words, the counterexample above is not robust. We will see in the next section that in one dimension the partially parallel algorithm indeed works in great generality.

The choice of $N = 3$ above is made for simplicity; one can easily see that for any odd value of $N \geq 3$ the model with periodic boundary conditions and equal and negative nearest neighbor couplings presents the type of behavior previously described.

If one generalizes the form of the energy (1), by introducing also external magnetic fields coupled to each spin, then one can construct another very simple example in which the limiting support is different from the set of global maxima of $H(\cdot)$. Consider just two spins, coupled ferromagnetically, but one subject to a positive and the other subject to a negative external field. Assume that the interaction and the external field are of the same strength. Again the process can escape with positive probability (uniformly in β) from the set of configurations that have maximum energy.

An apparent contradiction has to be solved: It is easy to construct counterexamples, although in many numerical experiments the limiting support is the set of maxima of H. Our aim is to clarify this apparent contradiction.

3. Limiting support for one dimensional spin glasses. Consider the one-dimensional version of the energy (1) on $\Lambda = \{1, \dots, N\}$:

$$H(\eta) = \sum_x J_{x,x+1}\eta(x)\eta(x+1) \tag{4}$$

with free or periodic boundary conditions (identify the site N with 0 and $N + 1$ with 1 in this latter case). In this section we will have to consider separately interactions of the following two types:

TYPE A. $J_{x,x+1} \geq 0$ for all $x \in \Lambda$.

TYPE B. The more general interactions: arbitrary $J_{x,x+1} \in \mathbb{R}$ for $x \in \Lambda$.

We will prove that with free or periodic boundary conditions the limiting support coincides with the set Ω^* of global maxima of the energy, provided an additional assumption is satisfied in Type B when periodic boundary conditions are used: Namely, that there is a unique site x for which $J_{x,x+1}$ is minimized. For free boundary conditions the result for interactions of Type A easily implies the same result for interactions of Type B. Periodic boundary conditions are more delicate: Recall the counterexample given in Section 2 with $J_{x,x+1} = -1$ and $N = 3$ (or any odd number), where the limiting support fails to be Ω^*. (Note that the additional assumption we will require excludes these cases.) The extra assumption above is automatically satisfied with probability one in case the $J_{x,x+1}$'s are drawn from independent absolutely continuous distributions.

Observe that the particular case in which $J_{x,x+1} = c > 0$ can be easily handled, since one can check that $I_{\tau,\infty}$ is the set of convex linear combinations of $\delta_{+\underline{1}}$ and $\delta_{-\underline{1}}$, the point masses concentrated on the configurations with all spins up or down. Hence the limiting support is equal to Ω^* by (3). However, in general, $I_{\tau,\infty}$ is a larger set, including point measures concentrated on "metastable configurations," that is, configurations which do not maximize the energy, but are such that any single flip of a spin strictly decreases the energy.

The first step in our approach to studying the Gibbs sampler for models of the form (4) consists in showing that they are *generalized voter models*. In order to make this statement precise we first observe that the probability measure $q_x(\eta, \cdot)$ can be obtained as follows. Define

$$\varepsilon_x = 2\frac{\exp(-\beta(|J_{x-1,x}| + |J_{x,x+1}|))}{\exp(-\beta(|J_{x-1,x}| + |J_{x,x+1}|)) + \exp(\beta(|J_{x-1,x}| + |J_{x,x+1}|))},$$

where $J_{0,1} = J_{N,N+1} = 0$ for free boundary conditions, and $J_{0,1} = J_{N,N+1} =: J_{N,1}$ for periodic boundary conditions. Let ρ_x be the number in $[0,1]$ which satisfies

$$\frac{\varepsilon_x}{2} + (1-\varepsilon_x)\rho_x = \frac{\exp(-\beta(|J_{x-1,x}| - |J_{x,x+1}|))}{\exp(-\beta(|J_{x-1,x}| - |J_{x,x+1}|)) + \exp(\beta(|J_{x-1,x}| - |J_{x,x+1}|))}$$

and define $\lambda_x = 1 - \rho_x$.

The measure $q_x(\eta, \cdot)$ gives positive mass only to configurations that differ from η at most at x, and here is how we can obtain the distribution of the spin at site x:

(i) With probability ε_x do the following: Assign with equal probabilities the value $+1$ or -1 to the spin at the site x.

(ii) With probability $1 - \varepsilon_x$ do the following: Assign to the spin at x the value $\eta(x+1)\operatorname{sgn}(J_{x,x+1})$ with probability ρ_x, or the value $\eta(x-1)\operatorname{sgn}(J_{x-1,x})$ with probability $\lambda_x = 1 - \rho_x$ [where $\operatorname{sgn}(z) = +1$ if $z \geq 0$ and $\operatorname{sgn}(z) = -1$ if $z < 0$].

Later we will interpret step (i) as noise with respect to (ii). The following statement is easy to prove by straightforward computations.

LEMMA 1. *The local characteristics* $q_x(\eta, \cdot)$ *on* $N(\eta, x) = \{\sigma: \sigma(y) = \eta(x), y \neq x\}$ *can be obtained by following steps* (i) *and* (ii).

REMARK. In case $J_{x-1,x} = 0$ and $J_{x,x+1} \neq 0$ we obtain $\lambda_x = 0$ and $\rho_x = 1$, and similarly in the other case. In particular free boundary conditions correspond to zero couplings at the boundaries and $\lambda_1 = \rho_N = 0$. In the case $J_{x-1,x} = J_{x,x+1} = 0$ we obtain $\varepsilon_x = 1$ so that λ_x and ρ_x are not defined and are not relevant.

Lemma 1 leads to a graphical construction of the various Gibbs sampler dynamics defined in Section 2. We refer the reader to Griffeath [8], Liggett [10] and Durrett [4] for reviews of this technique. We will represent the Gibbs sampler dynamics on $\{1, \dots, N\} \times \mathbb{Z}$, the space-time diagram. We first define random arrows between points in this space, then random marks on the points, corresponding to the activation of a site at a certain time and to the presence of noise. Next we will introduce paths in the space and rules that allow us to compute the value of the spin at a site at a certain time by following the path originating in it. Both the sequential and the partially parallel dynamics will be represented together, on the same diagram.

Arrows and marks. Draw an arrow from each point (x, t) to (i) $(x - 1, t - 1)$ with probability λ_x, or to (ii) $(x + 1, t - 1)$ with probability ρ_x, independently for each point. (See Figure 1.) To each arrow we attach a sign, defined as $\operatorname{sgn}(J_{x-1,x})$ and $\operatorname{sgn}(J_{x,x+1})$ for arrows of type (i) and (ii), respectively. These signs are necessary to deal with Type B interactions.

Next we mark each point (x, t) independently with an N (for noise) with probability ε_x. To each point (x, t) we associate a random variable $\alpha(x, t)$ taking values $+1$ and -1 with equal probabilities. These variables are used in order to assign values to the spins at the points where the noise is acting.

Mark each point independently with a U (for update) with probability τ. These marks are used by the partially parallel dynamics. We need different activation marks for the sequential Gibbs sampler: For each time t, choose a site uniformly in $\{(1, t), \dots, (N, t)\}$ and mark it with U', independently for different t.

On this probability space we construct both the partially parallel and the sequential Gibbs sampler. We denote by P the corresponding probability measure.

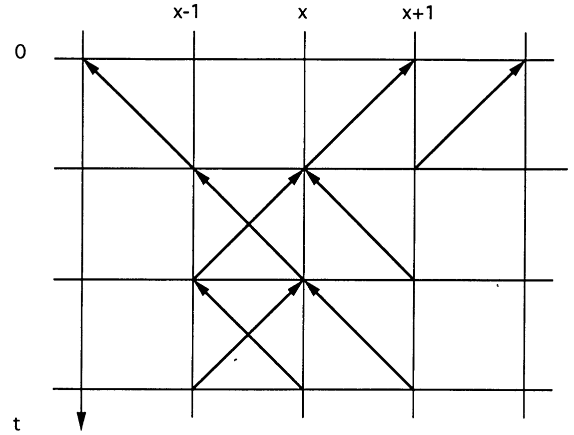

FIG. 1. *A realization of some arrows in the space time diagram.*

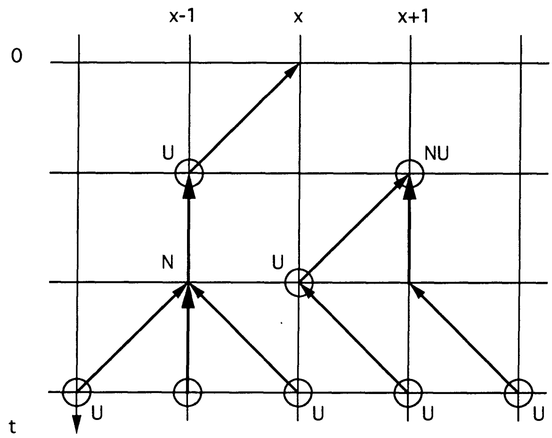

FIG. 2. *Some paths for the partially parallel Gibbs sampler.*

Paths. Starting from any point, paths will go backwards in time. For the partially parallel dynamics, the path starting from (x, t) is defined inductively as follows. When the path reaches a point (y, s) with $s < t$, then (1) if (y, s) is not marked with a U, then go to $(y, s - 1)$. Otherwise (2) if (y, s) is marked with N, then *stop* the path there. Otherwise (3) follow the arrow which starts at (y, s) [to $(y - 1, s - 1)$ or $(y + 1, s - 1)$].

For the sequential dynamics, the definition is analogous, but with U' replacing U. See Figures 2 and 3 for examples of paths for the partially parallel and the sequential Gibbs sampler, respectively. To facilitate the reading of the figures,

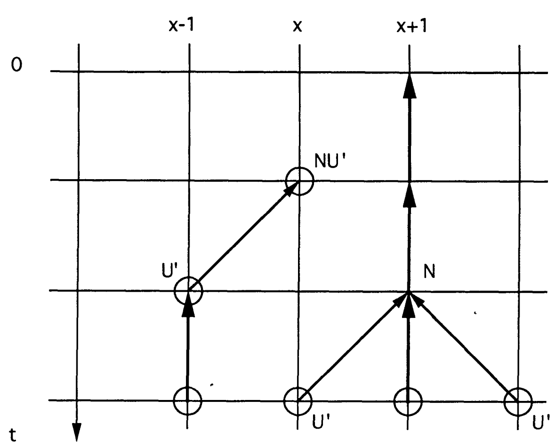

FIG. 3. *Some paths for the sequential Gibbs sampler.*

only some arrows and marks are displayed. We are ready to introduce the rules that allow us to identify the configuration at each time.

Given η_0, we compute $\eta_t(x)$ as follows:

1. If the path started at (x,t) is stopped at (y,s), with $s > 0$, then

$$\eta_t(x) = \kappa\alpha(y,s),$$

where κ is the product of the signs attached to the arrows used by the path.

Otherwise:

2. If the path started at (x,t) reaches $(z,0)$,

$$\eta_t(x) = \kappa\eta_0(z),$$

where κ is as above.

The graphical representation is now complete. Observe that for both dynamics the paths started at different points coalesce, in the sense that they stay together once they meet.

The crucial observation now is that in the case

$$\tau = \frac{1}{N},$$

the law of the path starting in a given point is the same for both the sequential and the partially parallel Gibbs sampler. This holds in spite of the fact that the joint law of two or more such paths is different for the two dynamics. This property allows us to draw conclusions about the partially parallel Gibbs sampler from our knowledge about the sequential one.

The invariant measure for the dynamics can be constructed as follows. Starting from $t = 0$ say, follow all paths from $(x,0)$, $x = 1,\dots,N$, back in (negative) time until each of them eventually stops. Assign to $\eta(x)$ the value $\kappa\alpha(y,-s)$ if the path from $(x,0)$ stops in $(y,-s)$ and κ is the product of the signs of the arrows followed along this path. One can easily see that $\{\eta(1),\dots,\eta(N)\}$ is distributed according to the invariant measure of the corresponding Gibbs sampler.

We are ready to state and prove our results, first for free boundary conditions and then for periodic ones.

THEOREM 2. *Consider interactions of Type B, with free boundary conditions. For every $\tau \in (0,1)$ the measure $\mu_{\tau,\infty}$ is concentrated on Ω^*.*

PROOF. We will exploit the property, proved by Trouvé [13], that $\mu_{\tau,\infty}$ exists and its support is independent of $\tau \in (0,1)$. Based on this we choose $\tau = 1/N$.

Consider first interactions of Type A. We can suppose that all $J_{x,x+1}$ are strictly positive, since otherwise the system will be broken into independent systems of the same type. Ω^* contains only the two configurations $-\underline{1}$ and $+\underline{1}$.

In order to prove certain facts about the graphical construction, it is convenient to consider the system with the special boundary condition in which the spin at the site 1 is frozen to $+1$, while at the other end of the lattice we keep the free boundary condition. In this case the corresponding Gibbs measure becomes concentrated on $+\underline{1}$, as $\beta \to \infty$. We can construct both the sequential and the partially parallel Gibbs sampler with this boundary condition by ignoring all marks N, U, U' that may appear at the points $(1, \cdot)$: When a path from (x, t) hits the site 1, it is stopped and $\eta_t(x)$ is set to the value $+1$. Define the events

$\mathscr{E}_p(x, y) =$ {the path of the partially parallel dynamics starting in $(x, 0)$ hits the site y before being stopped (by noise)},

and

$\mathscr{E}_s(x, y) =$ {the path of the sequential dynamics starting in $(x, 0)$ hits the site y before being stopped (by noise)}.

Considering now the sequential dynamics, one can conclude that for all $x \in \{1, \dots, N\}$,

$$\mu_\beta(\eta: \eta(x) = -1) \geq \tfrac{1}{2}\big(1 - P(\mathscr{E}_s(x, 1)\big).$$

Hence

$$\lim_{\beta \to \infty} P\big(\mathscr{E}_s(x, 1)\big) = 1.$$

Note that the event $\mathscr{E}_s(x, 1)$ is unaffected by the changes at the site 1 made above, since it does not depend on the value of the spin at this site nor on the presence of marks there. But since the definitions of the events $\mathscr{E}_s(x, y)$ and $\mathscr{E}_p(x, y)$ are based on a single path, these events have the same probability. Therefore for all $x \in \{1, \dots, N\}$,

(5) $$\lim_{\beta \to \infty} P\big(\mathscr{E}_p(x, 1)\big) = 1.$$

Similarly one can prove that for all $x \in \{1, \dots, N\}$,

(6) $$\lim_{\beta \to \infty} P\big(\mathscr{E}_p(x, N)\big) = 1.$$

Define the events

$\mathscr{F}(x, y) =$ {the paths of the partially parallel dynamics started at $(x, 0)$ and $(y, 0)$ coalesce before being stopped (by noise)}.

We want to prove that for all x, y,

(7) $$\lim_{\beta \to \infty} P(\mathscr{F}(x, y)) = 1.$$

This will imply that

$$\lim_{\beta \to \infty} \mu_{\tau,\beta}\{\eta\colon \eta(x) \text{ is constant and equal to } +1 \text{ or } -1\} = 1$$

and will finish the proof in this case.

From (5) and (6) it is clear that the paths started at $(x, 0)$ and $(y, 0)$ are likely to cross each other, but in principle they may jump over each other without coalescing. In fact if $\tau = 1$, this certainly occurs when $|x - y|$ is odd and we will exploit the fact that $\tau = 1/N < 1$. The idea is simply that (5) and (6) assure that the two paths cross each other many times before they are stopped by noise, by the strong Markov property of the hitting times of sites 1 or N. Each time the two paths cross, they have a chance to hit each other and coalesce. To be precise, suppose $x < y$ and define the stopping time (remember we are using the reversed time scale)

$T = \inf\{-t$: the paths started from $(x, 0)$ and $(y, 0)$ are respectively at time t in the sites z and $z + 1$, for some $z \in \{1, \dots, N - 1\}$, the points (z, t) and $(z + 1, t)$ are not marked with noise N, and there are arrows from (z, t) to $(z + 1, t - 1)$ and from $(z + 1, t)$ to $(z, t - 1)\}$.

It is clear that up to time T the paths did not cross each other, and with probability $2\tau(\tau - 1)$, independently of anything else, the two paths will hit each other at time $T + 1$. (See Figure 4.) By using now the strong Markov property and repeating this argument each time the two paths cross each other, one can prove (7) in a standard way.

We turn now to interactions of Type B. A simple transformation maps a model with such interactions into one with interactions of Type A. Indeed,

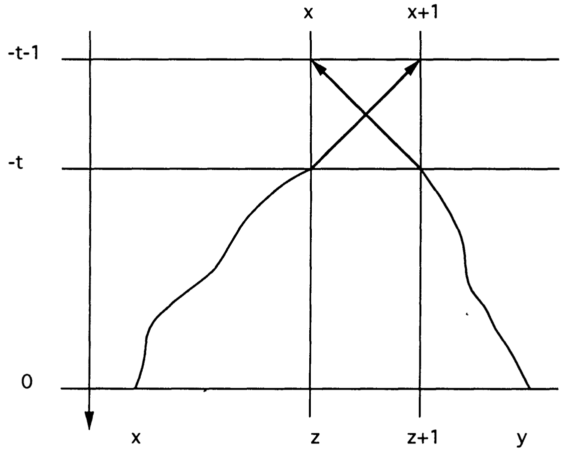

FIG. 4. *Crossing paths.*

consider the sequence

$$\zeta(1) = 1,\ \zeta(x+1) = \mathrm{sgn}(J_{x,x+1})\zeta(x) \quad \text{for } x > 0.$$

Let $\zeta\eta$ be defined as the sitewise product: $\zeta\eta(x) = \zeta(x)\eta(x)$. It is now easy to see that the states $\zeta(+\underline{1})$ and $\zeta(-\underline{1})$ are the global maxima for interactions of Type B and that the limiting support contains only these two configurations. □

Our next result concerns periodic boundary conditions.

THEOREM 3. *Consider interactions of Type A or B, with periodic boundary conditions. Assume in the latter case that there is only one site x for which* $|J_{x,x+1}|$ *is minimized. Then, for every* $\tau \in (0,1)$ *the measure* $\mu_{\tau,\infty}$ *is concentrated on* Ω^*.

PROOF. Consider first interactions of Type B. For the sake of notation assume that the smallest $|J_{x,x+1}|$ is $|J_{N,1}| (= |J_{0,1}| = |J_{N,N+1}|)$. We first identify the elements of Ω^*. The two states of maximum energy are obtained by ignoring the weakest coupling between N and 1, and satisfying all other couplings, in the following sense: For each pair $x < y$ define

$$\kappa(x,y) = \mathrm{sgn}(J_{x,x+1})\mathrm{sgn}(J_{x+1,x+2}) \cdots \mathrm{sgn}(J_{y-1,y})$$

and let $\kappa(x,x) = 1$ and $\kappa(x,y) = \kappa(y,x)$. Then the two states η^*, η^{**} with maximum energy are given by

$$\eta^*(x) = \kappa(1,x),$$
$$\eta^{**}(x) = -\kappa(1,x)$$

for $x \in \{1,\dots,N\}$.

Observe now that once two paths starting in $(x,0)$ and $(y,0)$ coalesce, the product of their signs becomes constant. So in case the paths coalesce before either one is stopped by noise, we have $\eta(x)\eta(y) =$ (the product of the signs of these two paths at the moment they coalesce). Also note that if these two paths coalesce without ever jumping from site 1 to site N or vice versa, then we get from the equality above

$$\eta(x)\eta(y) = \kappa(x,y).$$

Next we define an event which is similar to that one considered in the previous proof:

$\tilde{\mathscr{F}}(x,y) = \{$the paths of the partially parallel dynamics started at $(x,0)$ and $(y,0)$ coalesce before they are stopped (by noise), and before either one ever jumps from 1 to N or from N to 1$\}$.

As before we will be done once we prove that

$$\lim_{\beta\to\infty} P(\tilde{\mathscr{F}}(x,y)) = 1. \tag{8}$$

This can be proven similarly to the proof of (7). Since ε_x, ρ_x and λ_x are independent of the signs of the interactions $J_{x,x+1}$, we can assume with no loss of generality that the system is *frustrated*, that is, $\mathrm{sgn}(J_{N,1}) \neq \kappa(1, N)$.

Define the events

$\tilde{\mathscr{E}}_p(x, y) =$ {the path of the partially parallel dynamics starting in $(x, 0)$ hits the site y before it is stopped (by noise) and before it ever jumps from 1 to N or from N to 1}.

Consider now the boundary condition in which the spin at the site 1 is frozen to the value $+1$ (but N is still coupled to 1 by $J_{N,1}$). The measure $\mu_{\tau,\infty}$ is concentrated on η^*. Proceeding as in the previous proof, by using the sequential dynamics and observing that, for $\tau = 1/N$, single paths in both dynamics have the same law, we conclude that for all $x \in \{1, \dots, N\}$,

$$\lim_{\beta\to\infty} P\Big(\tilde{\mathscr{E}}_p(x,1)\Big) = 1. \tag{9}$$

Analogously,

$$\lim_{\beta\to\infty} P\Big(\tilde{\mathscr{E}}_p(x,N)\Big) = 1. \tag{10}$$

Now (8) follows from (9) and (10) as (7) followed from (5) and (6).

Finally we sketch the proof in the case of interactions of Type A. By the same argument used in the proof of Theorem 2, it follows that the path starting from any site is likely to hit the site 1 and the site N before it is stopped by noise. For periodic boundary conditions, this does not eliminate in principle the possibility that two paths, started at $(x, 0)$ and $(y, 0)$, never cross each other before they are stopped by noise. The reason for this is that they may rotate in the same direction on the circle $\{1, \dots, N\}$. For this to be the case, the paths starting from the sites 1 and N would have to be likely (as $\beta \to \infty$) to hit the sites N and 1, respectively, both going in the same direction, clockwise or counterclockwise. Now change the sign of the interaction $J_{N,1}$, in order to frustrate the system. As already observed, this does not affect the paths. Now use boundary conditions with the spin at site 1 or N frozen to the value $+1$. If the probability of the path from 1 to N going clockwise (respectively counterclockwise) tends to one as $\beta \to \infty$, then the probability of the path from N to 1 going counterclockwise (respectively clockwise) also tends to 1. And hence the paths cross each other, and the problem mentioned above does not occur. □

4. Limiting support for the two dimensional Ising model. In this section we consider the two-dimensional nearest neighbor ferromagnetic Ising model, and we show that for this model the limiting support coincides also with the set of global maxima of the energy. The proof is based on estimates of hitting times of the corresponding Markov chain and relies less on special tricks than the proofs presented in Section 3. For this reason we believe that the sort of arguments used here may turn out to be applicable also to other more interesting models.

THEOREM 4. *Consider the ferromagnetic Ising model on a finite square* Λ *contained in* $\mathbb{Z}^2$, *described by the usual nearest neighbor interaction*

$$H(\eta) = \frac{1}{2} \sum_{\langle x, y \rangle} \eta(x)\eta(y)$$

with periodic or free boundary conditions. For $0 < \tau < 1$,

$$\mu_{\tau,\infty} = \tfrac{1}{2}\delta_{+\underline{1}} + \tfrac{1}{2}\delta_{-\underline{1}}.$$

PROOF. We will present the proof in the case of periodic boundary conditions; the modifications needed to treat the case with free boundary conditions are straightforward and are omitted.

Let $A = \{+\underline{1}, -\underline{1}\} = \Omega^*$ be the set of maxima of H and B its complement. Because the partially parallel process is a renewal process with respect to visits to A,

$$\mu_{\tau,\infty}(A) = E(T_A)/(E(T_A) + E(T_B)),$$

where T_A and T_B are the times spent by the process in each one of these two sets during one renewal cycle. The expected time spent in A is inversely proportional to the probability of exiting A at each unit of time; this probability is proportional to $\exp(-4\beta)$. So the proof will be complete once we prove the following:

LEMMA 2. *Given* $\eta \in B$, *let* $(\eta_t)_{t \geq 0}$ *be the process started from* η *and set*

$$T^{\eta}(A) = \inf\{t \geq 0 : \eta_t \in A\}.$$

Then there is a constant c which does not depend on β *and* η, *but may depend on* $N := |\Lambda|^{1/2}$ *and* τ, *such that*

$$E(T^{\eta}(A)) \leq ce^{2\beta}.$$

In what follows all constants $c_1, c_2, \ldots$ will be finite and strictly positive and will not depend on β and η; they depend however on N and τ. (Note that the assumption $\tau < 1$ will be needed to guarantee that some of these c_i are strictly positive.)

PROOF OF LEMMA 2. Divide B into two sets C and $D = B \setminus C$, where C is the subset of configurations of B in which there is at least one site that has two neighbors in state $+1$ and two neighbors in state -1. It is easy to see that starting from any $\eta \in C$, the probability of the event Ω_1 that by time N^2 the system has hit $D \cup A$, is bounded from below by a constant $c_1 > 0$.

Starting from $\eta \in D$, we estimate now the probability of the event Ω_2, that by time $Ne^{2\beta}$ the system has hit either A or another configuration in D with more spins up than there are in η itself. This probability is bounded from below by the product of two terms: The first is the probability of the event Ω_3, that up to this time none of the $+1$ spins in η flipped. The second term is the conditional probability of the event Ω_4, given Ω_3, that before this time, N spins with value -1 did flip to $+1$ in successive time units, thus creating a new line of $+1$ spins adjacent to an already present one. Using $\lfloor \cdot \rfloor$ for integer part, we clearly have

$$P(\Omega_3) \geq \left(1 - c_2 N^2 e^{-2\beta}\right)^{\lfloor N e^{2\beta} \rfloor} \geq c_3$$

and

$$P(\Omega_4 | \Omega_3) \geq 1 - \left(1 - c_4 e^{-2\beta}\right)^{\lfloor e^{2\beta} \rfloor} \geq c_5.$$

To derive the second estimate, observe that there are $\lfloor e^{2\beta} \rfloor$ independent chances for N successive -1 spins along a line to flip. In each attempt the probability of success is at least $c_4 e^{-2\beta}$, since, after the first one flips, all the other can flip easily.

Therefore we have

$$P(\Omega_2) \geq P(\Omega_3, \Omega_4) \geq c_6.$$

One can now repeatedly use the estimate above, when the system first enters D and each time it reaches a new configuration in D with more $+1$ spins than the last one visited previously in D. Since the system will have to hit A before visiting N different configurations in D in such a manner, we obtain for all $\eta \in B$:

$$P(T^{\eta}(A) \leq N^2 e^{2\beta}) \geq P(\Omega_1)\left(P(\Omega_2)\right)^N \geq c_1 (c_6)^N = c_7.$$

Therefore for all $k = 1, 2, \dots,$

$$P(T^{\eta}(A) \geq k N^2 e^{2\beta}) \leq (1 - c_7)^k$$

and

$$E(T^{\eta}(A)) \leq N^2 e^{2\beta} \sum_{k=0}^{\infty} P\left(T^{\eta}(A)/(N^2 e^{2\beta}) \geq k\right) = c e^{2\beta}.$$

This completes the proof of the lemma and Theorem 3 follows from it. □

We finally observe that also if the interactions between nearest neighbors are not strictly identical, but are close enough to a common positive value, the proof given above can be applied to conclude that the partially parallel algorithm works.

Acknowledgments. Thanks are due to Alain Trouvé and Laurent Younes for introducing us to the problem and to Tom Liggett for a helpful discussion that allowed us to avoid a long proof.

REFERENCES

[1] AARTS, E. H. L. and VAN LAARHOVEN, P. J. M. (1987). *Simulated Annealing: Theory and Applications*. Reidel, Dordrecht.

[2] AZENCOTT, R. (1992). Boltzmann machines: High order interactions and synchronous learning. *Lecture Notes in Statist.* Springer, Berlin. To appear.

[3] DAWSON, D. A. (1975). Synchronous and asynchronous reversible Markov systems. *Canad. Math. Bull.* **17** 633–649.

[4] DURRETT, R. (1988). *Lecture Notes on Particle Systems and Percolation*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

[5] GAUDRON, I. and TROUVÉ, A. (1990). Parallelization of simulated annealing for spin glasses energies. Preprint. Univ. Paris-Sud.

[6] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Trans. PAMI* **6** 721–741.

[7] GEMAN, D. (1990). Random fields and inverse problems in imaging. *Ecole d'Eté de Saint-Flour XVIII. Lecture Notes in Math.* **1427** 117–193. Springer, Berlin.

[8] GRIFFEATH, D. (1979). *Additive and Cancellative Interacting Particle Systems. Lecture Notes in Math.* **724**. Springer, Berlin.

[9] HWANG, C. R. and SHEU, S.-J. (1987). Large time behavior of perturbated diffusion Markov processes with applications. Parts I, II and III. Technical reports of the Institute of Mathematics, Academia Sinica, Taiwan.

[10] LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, Berlin.

[11] SOKAL, A. (1989). Monte Carlo methods in statistical mechanics: Foundations and new algorithms. Technical report, Univ. de Lausanne.

[12] TROUVÉ, A. (1988). Problèmes de convergence et d'ergodicité pour les algorithmes de recuit parallélisés. *C. R. Acad. Sci. Paris Ser. I* **307** 161–164.

[13] TROUVÉ, A. (1990). Parallel simulated annealing and effective optimization: study of the limit measure. Preprint. Univ. Paris-Sud.

PABLO A. FERRARI
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
CX. POSTAL 20570
01498 SÃO PAULO
BRASIL

ARNOLDO FRIGESSI
ISTITUTO PER LE APPLICAZIONI
DEL CALCOLO DEL CNR
VIALE DEL POLICLINICO 137
00161 ROMA
ITALIA

ROBERTO H. SCHONMANN
DEPARTMENT OF MATHEMATICS
UCLA
LOS ANGELES, CALIFORNIA 90024