On the Existence and Non-Existence of Finitary Codings for a Class of Random Fields

J. van den Berg
CWI
P.O. Box 94079, 1090 GB Amsterdam
The Netherlands

J.E. Steif
Department of Mathematics
Chalmers University of Technology, S-41296 Gothenburg
Sweden

Abstract

We study the existence of finitary codings (also called finitary homomorphisms or finitary factor maps) from a finite-valued i.i.d. process to certain random fields. For Markov random fields we show, using ideas of Marton and Shields, that the presence of a phase transition is an obstruction for the existence of the above coding: this yields a large class of Bernoulli shifts for which no such coding exists.

Conversely, we show that for the stationary distribution of a monotone exponentially ergodic probabilistic cellular automaton such a coding does exist. The construction of the coding is partially inspired by the Propp-Wilson algorithm for exact simulation.

In particular, combining our results with a theorem of Martinelli and Olivieri, we obtain the fact that for the plus state for the ferromagnetic Ising model on $\mathbf{Z}^d$, $d \geq 2$, there is such a coding when the interaction parameter is below its critical value and there is no such coding when the interaction parameter is above its critical value.

1991 Mathematics Subject Classification: 28D99 60K35 82B20 82B26.
Key words and Phrases: Random fields, Phase transitions, Finitary coding.
Note: Work by the first author carried out under project PNA3.1 "Probability". A slightly different version of this paper has been accepted for publication in the Annals of Probability.

1 Introduction

One of the main motivations for our paper is the following (all definitions will be given later): In [34] (see [1] for a published version), it is proved that the plus state for the Ising model (with 0 external field) is a Bernoulli shift (i.e., is isomorphic to an i.i.d. process) below, above and at the critical value of the interaction parameter. Therefore, although there are important differences in the behavior of the plus state above and below the critical value, these differences are not reflected in the notion of isomorphism. It turns out however that these differences are reflected if one considers the notion of *finitary* mappings instead. The following theorem (restated and proved in Section 4) is a particular case of the general results obtained in our paper.

Theorem 1.1 *There does not exist a finitary factor map from any finite-valued i.i.d. process to the plus state for the Ising model above the critical interaction parameter. However, there does exist a finitary factor map from a finite-valued i.i.d. process to the plus state for the Ising model below the critical interaction parameter.*

In fact, one direction of Theorem 1.1 will follow from the following more general theorem (restated and explained in Section 3).

Theorem 1.2 *The limit distribution μ of a monotone, exponentially ergodic probabilistic cellular automaton is a finitary factor of a finite-valued i.i.d. process.*

We will now give background and the necessary definitions. *Throughout this paper, all stationary processes and stationary random fields will be assumed to be finite-valued unless otherwise stated, $\|\ \|$ will denote the L_1 norm on $\mathbf{Z}^d$ given by $\|x\| = |x_1| + |x_2| + \ldots + |x_d|$, and we will often write $[a, b]$ for $[a, b] \cap \mathbf{Z}$ and $[a, b]^d$ for $[a, b]^d \cap \mathbf{Z}^d$.*

In [33], D. Ornstein proved the celebrated isomorphism theorem for Bernoulli shifts. This states that if $\{X_n\}_{n\in\mathbf{Z}}$ and $\{Y_n\}_{n\in\mathbf{Z}}$ are stationary processes consisting of independent and identically distributed (i.i.d.) random variables with equal entropy, then there exists a stationary a.e. invertible mapping from the first process to the second. More specifically, if μ is a probability measure on $A^{\mathbf{Z}}$ (A a finite set) which is a product measure with all the same marginals and if ν is a probability measure on $B^{\mathbf{Z}}$ (B a finite set) which is a product measure with all the same marginals and if

$-\sum_{i\in A} p_i \log p_i = -\sum_{i\in B} q_i \log q_i$, where $(p_i, i \in A)$ and $(q_i, i \in B)$ are the marginals of μ and ν respectively, then there exists an invertible measure preserving map from $(A^{\mathbf{Z}}, \mu)$ to $(B^{\mathbf{Z}}, \nu)$ which is defined a.e. and which commutes with shifts. When μ and ν are arbitrary probability measures on $A^{\mathbf{Z}}$ and $B^{\mathbf{Z}}$ respectively which are each invariant under the natural $\mathbf{Z}$–action, then a mapping with the above property is called an *isomorphism* and the two processes are then called *isomorphic*. If we drop the invertibility assumption, we call the mapping a *factor map* or *homomorphism* and say that the second process is a *factor* of the first. Often, instead of 'factor map' or 'homomorphism' the shorter term 'coding' is used in the literature (although some authors use this only for isomorphisms), and we will use this in most of the paper.

The σ–algebra involved above is the completed Borel σ–algebra with respect to the product topology. Actually, it is not necessary that A and B are finite sets and the above result holds more generally (see [33]). From this work, a number of properties of stationary processes emerged which implied that a given process is isomorphic to an i.i.d. process. Processes which are isomorphic to i.i.d. processes are called *Bernoulli shifts*.

It was also proved by Ornstein (see [33]) that a factor of an i.i.d. process is a Bernoulli shift.

Prior to [33], isomorphisms between certain classes of i.i.d. processes were obtained (see [5] and [31]). These mappings had the advantage of being finitary. A coding is called *finitary* if it is continuous after removing some set of measure 0. There is another more natural equivalent definition of finitary in this context (which also explains the word finitary). To describe this, if $z \in A^{\mathbf{Z}}$ and $q \leq r$ are integers, we let $z[q,r]$ denote $(z(q), z(q+1), \ldots, z(r))$. With this notation, ϕ is a finitary coding if and only if there exists a set $\mathcal{N} \subseteq A^{\mathbf{Z}}$ of μ–measure 0 such that for all $x \in A^{\mathbf{Z}} \backslash \mathcal{N}$, there exist integers $q \leq r$ (depending on x) so that if $y \in A^{\mathbf{Z}} \backslash \mathcal{N}$ and $y[q,r] = x[q,r]$, then $\phi(y)(0) = \phi(x)(0)$. In words, after a long enough *finite* subsequence of the x sequence is revealed, we know the 0th coordinate of $\phi(x)$. In [21], it is proved that there exists a finitary coding from any i.i.d. process onto any other i.i.d. process with strictly lower entropy. One of the ideas of this approach came from [32] where finitary isomorphisms were obtained between special Markov chains and i.i.d. processes. The work in [21] was extended in [4] to the case of general finite state mixing Markov chains. At the same time, it was proved in [22] that for any two i.i.d. processes with the same entropy, there exists a finitary isomorphism between them whose inverse is also finitary. Finally, after this, it was proved (see [23]) that for any two finite state Markov chains with the same entropy and period, there exists a finitary isomorphism between them whose inverse is also finitary. We mention that it was proved much earlier (see [13]) that a finite state mixing Markov chain is isomorphic (not necessarily finitarily) to an i.i.d. process.

For $d \geq 2$, one considers probability measures μ and ν on $A^{\mathbf{Z}^d}$ and $B^{\mathbf{Z}^d}$ respectively which are each invariant under the natural $\mathbf{Z}^d$–action. One calls such objects stationary random fields. An invertible measure preserving map from $(A^{\mathbf{Z}^d}, \mu)$ to $(B^{\mathbf{Z}^d}, \nu)$ which is defined a.e. and which commutes with shifts in the d directions is also called an *isomorphism* and the two processes are then also called *isomorphic*. The notions of (finitary) coding and Bernoulli shift extend immediately to $d \geq 2$ dimensions. The Ornstein isomorphism theorem also extends to $d \geq 2$ dimensions (see [20]) and in fact much further to amenable groups (see [35]). It is also mentioned in [20] that the theorem by Ornstein mentioned earlier that a factor of an i.i.d. process is a Bernoulli shift also extends to random fields.

Several results in this paper are about Markov random fields which, for completeness, we give the definition of. Let B_n denote $[-n, n]^d$ and for $U \subseteq \mathbf{Z}^d$, let $\partial(U)$ denote the boundary $\{x \in \mathbf{Z}^d \backslash U : \exists y \in U \text{ with } \|x - y\| = 1\}$. Further, if $U \subseteq \mathbf{Z}^d$, we use the notation $X_U = \{X_x\}_{x \in U}$.

Definition 1.3 *A stationary process* $\{X_x\}_{x \in \mathbf{Z}^d}$ *is called a* Markov random field *if, for all finite subsets* $B \subseteq \mathbf{Z}^d$*, the conditional distribution of* X_B *given* X_{B^c} *is the same as the conditional distribution of* X_B *given* $X_{\partial(B)}$.

While it is known that finite state mixing Markov chains in 1 dimension are isomorphic to i.i.d. processes, this is not so in higher dimensions. See [39] for examples of mixing (k-step) Markov random fields which are not K (the definition of K is given below) and hence are not Bernoulli shifts as well as a number of other interesting examples. Here, we also mention the recent book [37] where random fields which arise in an algebraic context are studied and where their dynamical properties are given certain algebraic characterizations. We finally mention that a Markov random field which is K but not a Bernoulli shift was recently constructed in [16], thereby giving a counterexample to a previous conjecture. The definition of K in general is slightly complicated but for Markov random fields it is shown (in [17]) to be equivalent to the property that the (full) tail σ–algebra, $\cap_{n \geq 1} \sigma(X_i, i \notin B_n)$, is trivial. See [17] for other related results concerning the K property and Bernoulli shifts for Markov random fields and [18] for extensions to Gibbs states.

As far as extending the results in [4] and [23] to higher dimensional Markov random fields, it was proved in [19] that there exists a finitary coding from any ergodic Markov random field onto any i.i.d. random field of strictly lower entropy. It is mentioned there that it is not known when a Markov random field is a Bernoulli shift and so one should not necessarily hope to prove that there exists a finitary coding from any i.i.d. random field onto any mixing Markov random field of strictly lower entropy.

It turns out, as we show in Section 2, that there are even Markov random fields which *are* Bernoulli shifts but *are not* a finitary factor of any i.i.d. random field. In fact, there is a fundamental obstruction to the existence of a finitary coding from an i.i.d.. random field onto a given random field, and there is a large number of Markov random fields which are Bernoulli shifts and possess this obstruction. Some of these are even measures of maximal entropy for nearest neighbor subshifts of finite type (see [6], [7] and [8]). Conversely, we show in Section 3 how, for certain random fields, a finitary coding from an i.i.d. process *can* be constructed. In Section 4 we treat, as an important special case, the ferromagnetic Ising model on $\mathbf{Z}^d$, and restate and prove Theorem 1.1.

2 A fundamental obstruction for finitary coding

In this section we show that many Markov random fields are not a finitary factor of an i.i.d. random field. The main result in this section is the following.

Theorem 2.1 *Let ν be an ergodic Markov random field all of whose cylinder sets have positive probability and with the property that there exists another (different) ergodic Markov random field ν' which has the same conditional probabilities as ν (i.e., for all finite sets $B \subseteq \mathbf{Z}^d$, the ν–conditional distribution of X_B given $X_{\partial B}$ is the same as the ν'–conditional distribution of X_B given $X_{\partial B}$). Then there does not exist a finite-valued i.i.d. random field μ and a finitary coding from μ onto ν.*

Remarks:
(a). This result can be extended to so-called infinite range Gibbs states with essentially the same proof.
(b). An obstruction for finitary codings between infinite state Markov chains due to M. Smorodinsky is given in [26]. This obstruction arises from the fact

that certain waiting times between states have tails which are not exponential. In our case, this latter behavior is not present and the nature of the obstruction is completely different.
(c). Before proving this theorem, we mention that there is a large class of Markov random fields which satisfy the assumptions of the theorem and which are Bernoulli shifts. The above theorem tells us that these random fields provide examples of Bernoulli shifts for which no i.i.d. process can be mapped onto them in a finitary fashion. An example of such a field is the "plus state" for the Ising model with sufficiently large interaction parameter (see Section 4). In Section 4, it is explained that this example satisfies the assumptions of Theorem 2.1 while the fact that it is a Bernoulli shift is proved in [34]. (Whenever we refer to [34] in this paper, if one wants to see published work on the same topic, one can refer to [1] which extends the work in [34] to an amenable group setting.) The above theorem also holds in many situations where one does not need to assume that all the cylinder sets have positive probability. For example, by modifying the proof below, one can obtain the same conclusion for some of the measures of maximal entropy in [6], [7] and [8].
(d). In the proof we give below, the reader may not see "what is really going on". For this reason, we first give a quick discussion explaining to some extent what is going on. There is an important property called the *blowing–up property* which says more or less that any collection of configurations on a large finite box which has a total measure which is not too exponentially small in the volume of the box has the property that most configurations are close to it in the Hamming metric. (The blowing–up property is related to the notion of *concentration of measure*, see [41].) A consequence of this blowing–up property is that the mean ergodic theorem holds at an exponential rate. Since i.i.d. processes have the blowing–up property (as mentioned in the proof), and finitary codings preserve this property (as also mentioned in the proof), any Markov random field which is a finitary factor of an i.i.d. process must have the mean ergodic theorem holding at an exponential rate. However, when an ergodic Markov random field is not the unique Markov random field with its conditional probabilities, this usually results in the mean ergodic theorem holding at a subexponential rate and a typical scenario is as follows. μ and ν are distinct ergodic Markov random fields with the same conditional probabilities but have different means. Roughly speaking, there is a certain boundary condition which you can place on $\partial(B_n)$ which has positive μ measure but such that conditioned on this boundary condition, what is inside looks like ν. The μ–probability that this particular

boundary condition arises is at least $e^{-cn^{d-1}}$ for some constant $c < \infty$ independent of n. However, if this boundary condition occurs, then with high (conditional) probability the average in the box will be the ν mean which is a fixed distance away from the μ mean. Therefore in this scenario, the mean ergodic theorem occurs at a subexponential rate.

In [30], the *blowing–up property* for a stationary ergodic process is discussed (see also Section 1.5 in [9] and references there for background). This definition, which we will need here, immediately extends to random fields and is the following.

Definition 2.2 *An ergodic stationary random field taking values in the set A and indexed by $\mathbf{Z}^d$ has the* blowing-up property *if given $\epsilon > 0$, there exists $\delta > 0$ and an N such that for all $n \geq N$, we have that if $C \subseteq A^{[-n,n]^d}$ with $P(C) \geq 2^{-(2n+1)^d \delta}$, then $P([C]_\epsilon) \geq 1 - \epsilon$ where $[C]_\epsilon$ is the set of all configurations $(a_i,\, i \in [-n,n]^d)$ for which there exists $(c_i,\, i \in [-n,n]^d) \in C$ with*

$$\frac{1}{(2n+1)^d} \sum_{i \in [-n,n]^d} I_{\{a_i \neq c_i\}} < \epsilon$$

i.e., $[C]_\epsilon$ is the ϵ–neighborhood of C in the $\overline{\mathrm{d}}$–metric.

Proof of Theorem 2.1: It is proved in [30] that finitary codings preserve the blowing–up property in 1-d. The proof of this result goes through step by step to $d \geq 2$ dimensions. It is proved in [9] that a 1-d i.i.d. process satisfies the blowing–up property. This result for 1-d immediately yields the same fact for higher dimensions. Therefore, in order to show that there does not exist an i.i.d. random field μ and a finitary coding from μ onto ν, we need only show that ν does not have the blowing–up property.

Next, it follows immediately from Theorem 1.1 in [30] that if a 1-d process μ has the blowing–up property, then given any other (different) ergodic process ν, the *lower divergence rate* of ν with respect to μ is positive where the lower divergence rate, also known as relative entropy, of ν with respect to μ for $\mathbf{Z}^d$ processes is defined as

$$\liminf_{n\to\infty} \frac{1}{(2n+1)^d} \sum_{a \in A^{[-n,n]^d}} \nu(a) \log(\frac{\nu(a)}{\mu(a)}).$$

The proof of this result easily extends to $d \geq 2$ dimensions. Finally, it is known (see [14], p. 322–323) that for any two Markov random fields all of whose cylinder sets have positive probability and which have the same conditional probabilities, the lower divergence rate of one with respect to the other is 0. Applying these facts to ν and ν', we conclude that ν does not have the blowing–up property. $\Box$

Remark(e): K. Marton has explained to us that the proof that a finitary factor of an i.i.d. process has the ergodic theorem occurring at an exponential rate for all functions, is simpler than the proof that finitary codings preserve the blowing-up property. As it is known that for any ν satisfying the assumptions of Theorem 2.1 the ergodic theorem cannot occur at an exponential rate for all functions, this would yield a simpler proof of Theorem 2.1. However, since the blowing-up property is stronger, it gives rise to a more powerful recipe for determining that a random field is not a finitary factor of an i.i.d. process. Moreover, the higher-dimensional generalizations above of results concerning the blowing–up property, are also useful outside the scope of this paper.

Remark(f): Finally, we mention that there are nontrivial Markov random fields in higher dimensions which have the blowing–up property. Theorem 2 in [29] easily extends to $d \geq 2$ dimensions. Using this corollary together with the arguments in [34], it follows that when there is a unique Markov random field for the Ising model, it has the blowing–up property. Therefore, the proof of Theorem 2.1 does not exclude the possibility that there exists a finitary coding from some i.i.d. process onto this Markov random field, and in Section 4 we show that this is, except possibly at the critical point, indeed the case.

3 Finitary codings for the limit distributions of exponentially ergodic probabilistic cellular automata

Let S be a finite set; this will be our single-site state space. We assign a linear order $\leq$ to S and denote the maximal and minimal element of S by $+$ and $-$ respectively. Let $\Omega = S^{\mathbf{Z}^d}$. With abuse of notation, $+$ and $-$ will also be used to denote the maximal and minimal element of S^V with the induced partial order (given by $(a_i, i \in V) \leq (b_i, i \in V)$ if $a_i \leq b_i$ for all $i \in V$) when $V \subseteq \mathbf{Z}^d$. We consider certain time evolutions on Ω. As in Section 2, if $\omega = (\omega_i, i \in \mathbf{Z}^d) \in \Omega$ and $V \subseteq \mathbf{Z}^d$, then ω_V is the 'restriction of ω to V', i.e., $\omega_V = (\omega_i, i \in V)$. Further, if μ is a probability distribution on

Ω (with the natural σ-field), then μ_V denotes the 'restriction of μ to V', i.e., $\mu_V(\cdot) = \mu(\omega_V \in \cdot)$. If μ and ν are two probability distributions on a finite set F, the variational distance of μ and ν (defined as $1/2 \sum_{x \in F} |\mu(x) - \nu(x)|$) is denoted by $d_v(\mu, \nu)$. The time evolutions we consider on Ω correspond with so-called *probabilistic cellular automata* (PCA's). To describe them in a way suitable for future use, let $W_{i,t}, i \in \mathbf{Z}^d, t \in \mathbf{N}$, be i.i.d. random variables taking values in a finite set A. Let, for $i \in \mathbf{Z}^d$, N_i denote the set of vertices at (lattice) distance ≤ 1 from i. Consider a function $f : S^{N_O} \times A^{N_O} \to S$. Define, for each $i \in \mathbf{Z}^d$, $f_i : S^{N_i} \times A^{N_i} \to S$ by $f_i(s, a) = f(s - i, a - i)$, where $s - i \in S^{N_O}$ is defined by $(s - i)_j = s_{i+j}$, $j \in N_O$, and $a - i \in A^{N_O}$ is defined similarly. The time evolution $\sigma(\omega, t), t = 0, 1, \cdots$, starting from an initial configuration ω is now described by

$$\begin{aligned} \sigma(\omega, 0) &= \omega. \\ \sigma_i(\omega, t+1) &= f_i((\sigma_j(\omega, t)\,, j \in N_i), (W_{j,t}\,, j \in N_i)), \;\; i \in \mathbf{Z}^d. \end{aligned} \tag{1}$$

Since the $W_{i,j}$'s are random, the $\sigma_i(\omega, t)$'s are also random.

Remark(g): If f does not depend on the A-variables, we have a deterministic cellular automaton. PCAs are usually defined somewhat differently from the above, namely with f being a function $S^{N_0} \times A \to S$. The interpretation is then that at each time the value of each vertex is replaced by a new (random) value, whose distribution depends on the current local configuration, and that, conditioned on $\sigma(\omega, t)$ the $\sigma_i(\omega, t + 1)$, $i \in \mathbf{Z}^d$ are *independent*. With the application in Section 4 to the Ising model in mind, we prefer the slightly more general setup which allows local conditional dependence.

Definition 3.1 *We say that a PCA is* monotone *if the function f above is monotone in its first argument (i.e., if for each α, $\beta \in S^{N_0}$ with $\alpha \geq \beta$, and each $a \in A^{N_O}$, $f(\alpha, a) \geq f(\beta, a)$).*

Let $\mu(\omega, t)$ denote the distribution of the configuration at time t when we start with configuration ω.

Definition 3.2 *We say that a PCA is* ergodic *if there exists a distribution μ on Ω such that for all $\omega \in \Omega$, $\mu(\omega, t)$ converges (weakly) to μ as $t \to \infty$.*

(Note that the word "ergodic" here has a different meaning from that in Sections 1 and 2, but it should always be clear from the context which is meant.) If the system is monotone, this is (as is well-known and can be easily checked by standard coupling arguments) equivalent to saying that

$$\lim_{t\to\infty} P(\sigma_O(-,t) \neq \sigma_O(+,t)) = 0.$$

(Note that by the construction of the PCA, the evolutions starting from different configurations are coupled and so this last probability makes sense.)

Definition 3.3 *We say that a monotone PCA is* exponentially ergodic *if there exists $C, \lambda > 0$ such that*

$$P(\sigma_O(-,t) \neq \sigma_O(+,t)) \leq C\exp(-\lambda t)$$

for all t.

The main result of this section (which is Theorem 1.2 in the introduction) is the following.

Theorem 3.4 *The limit distribution μ of a monotone, exponentially ergodic PCA (as defined above) is a finitary factor of a finite-valued i.i.d. process.*

Proof: The first part of the proof gives an algorithm for *exact simulation* of μ, i.e. a randomized algorithm which assigns to each vertex, one by one, a value in S, such that the resulting configuration has distribution μ. This is a modification of the popular (finite-volume) Propp-Wilson algorithm (see [36]) and is used as a starting point for the construction of a finitary coding.

As to the simulation procedure, it is convenient (and essential in the Propp-Wilson method) to extend time to the negative integers, so we now have $(W_{i,t}, i \in \mathbf{Z}^d, t \in \mathbf{Z})$. Completely analogous to what we did before, we can then consider, for each $t_1 < t_2$, the configuration we have at time t_2 when we start with configuration ω at time t_1. We denote this by $\Phi_{t_1}^{t_2}(\omega)$. (This depends of course on the $W_{i,t}$'s but we omit these from our notation.) Clearly, $\Phi_{t_1}^{t_2}(\omega)$ has distribution $\mu(\omega, t_2 - t_1)$ defined before. The idea is to start, for each $i \in \mathbf{Z}^d$, so far "backwards in time" that the spin value at site i at time 0 is the same for all starting configurations. More formally, define

$$\tau_i = \min\{t : (\Phi_{-t}^0(+))_i = (\Phi_{-t}^0(-))_i\}.$$

Clearly, if such a τ_i exists, then (using monotonicity) $\Phi_{-t}^0(\omega) = \Phi_{-t}^0(\omega')$ for all ω and $\omega' \in \Omega$ and $t \geq \tau_i$. In other words, if we start at or before time $-\tau_i$, then the value at vertex i at time 0 no longer depends on the starting configuration (or on the starting time, as long as it is smaller than or equal to $-\tau_i$). We denote this value by σ_i^*. So, formally, $\sigma_i^* = (\Phi^0_{-\tau_i}(+))_i$. Of course we have to show that τ_i exists (i.e., is finite a.s.). Once we have done this, and, moreover, have shown that $\sigma^* \equiv (\sigma_i^*, i \in \mathbf{Z}^d)$ has the desired distribution μ, then it is clear what the simulation algorithm is: Determine for each $i \in \mathbf{Z}^d$, one by one, the value σ_i^* as follows. Check, for larger and larger values of t, if $(\Phi^0_{-t}(+))_i$ equals $(\Phi^0_{-t}(-))_i$. If this is the case, assign their common value to σ_i^*.

Remark(h): Note that for each t the above mentioned check is a finite task, since it involves only those $W_{j,t'}$, $j \in \mathbf{Z}^d$, $t' \in \{-t, \cdots, -1\}$ with $\| j-i \| \leq |t'|$ where $\| \ \|$ is the L_1 norm on $\mathbf{Z}^d$. During the algorithm a $W_{j,t}$ is generated the first time it is needed in a calculation and, which is important, of course keeps its value during the remainder of the algorithm. We soon come back to this notion of 'being needed', which will be essential in our construction of a finitary coding.

Lemma 3.5 *If the PCA is monotone and ergodic, then each τ_i is finite a.s.*

Proof of Lemma 3.5: Let μ be the limit distribution of the PCA. Using the monotonicity of the PCA,

$$\begin{aligned} P(\tau_i > t) &= P[(\Phi^0_{-t}(+))_i \neq (\Phi^0_{-t}(-))_i] \\ &= P(\sigma_O(-,t) \neq \sigma_O(+,t)) \end{aligned} \tag{2}$$

which (by the assumption that the system is ergodic) goes to 0 as t goes to ∞. □

Lemma 3.6 *The random configuration σ^*, defined above, has distribution μ.*

Proof of Lemma 3.6: Let Λ be a finite subset of $\mathbf{Z}^d$, and let σ be a random configuration on $\mathbf{Z}^d$, drawn according to distribution μ. Since μ is invariant under the dynamics, we have for every t, that $\Phi^0_{-t}(\sigma)$ has distribution μ. In particular we have $(\Phi^0_{-t}(\sigma))_\Lambda$ has distribution μ_Λ for every t. However, by Lemma 3.5, $(\Phi^0_{-t}(\sigma))_\Lambda = (\sigma^*)_\Lambda$ for all sufficiently large t. Hence σ^*_Λ has distribution μ_Λ. This holds for every finite $\Lambda \subseteq \mathbf{Z}^d$, which completes the proof. □

Remark(i): Letting $Z_i = \{W_{i,t}, t \in \mathbf{Z}\}$, we have that $\{Z_i\}_{i\in\mathbf{Z}^d}$ is an i.i.d. process and that μ is a finitary coding of it. The point however is that $\{Z_i\}_{i\in\mathbf{Z}^d}$ is not a finite-valued process. The idea is now to modify the process $\{Z_i\}_{i\in\mathbf{Z}^d}$ and the simulation algorithm so that we can obtain a finitary coding from a *finite-valued* i.i.d. process to μ.

The proofs of Lemmas 3.5 and 3.6 have not used the exponential convergence. However this exponential convergence yields the fact that (as we point out more precisely below) for each i the *expected* number of $t \in \{\cdots, -2, -1\}$ for which there exists a $j \in \mathbf{Z}^d$ such that $W_{i,t}$ is "needed for the computation of σ_j^*" is finite.

Remark(j): It is not difficult to show that this exponentiality even implies that, for the above described simulation procedure, the expected amount of computational work to generate the spin values of n vertices is bounded by a constant times n. Although this is not essential for our purpose, it is interesting from a simulation point of view, since it says that, in some sense, this algorithm has linear rate, while it is clear that it is impossible to do essentially better than linear.

Consider the random variables $\{W_{i,t}, i \in \mathbf{Z}^d, t \leq -1\}$. Informally, we say that the position (i,t) $(i \in \mathbf{Z}^d, t \leq -1)$ in the space time diagram is *needed* if, for some $j \in \mathbf{Z}^d$, it is 'involved' in the evaluation of σ_j^*. More precisely, we define that (i,t) is needed if there is some $j \in \mathbf{Z}^d$ such that $\tau_j \geq |t|$ and $\| j - i \| \leq |t|$; we let $n(i,t)$ denote the indicator function of this event. If (i,t) is not needed and the W variable is changed at (i,t), then clearly σ^* is unaffected. Next, the expected number of t such that $(0,t)$ is needed equals

$$\sum_{t=-\infty}^{-1} P(\cup_{j:\|j\|\leq|t|}\{\tau_j \geq |t|\}) \leq \sum_{t=-\infty}^{-1} |\{j :\| j \|\leq |t|\}| P(\tau_0 \geq |t|) < \infty,$$

by (2) and the exponentially ergodic assumption.

Let M be any integer larger than this expectation and

$$L := M - E[\sum_{t=-\infty}^{-1} n(0,t)] > 0$$

be the difference. For $j \in \mathbf{Z}^d$, let $n(j) = \sum_{t=-\infty}^{-(M+1)} n(j,t)$ and $s(j) = M - \sum_{t=-M}^{-1} n(j,t)$. For $i \in \mathbf{Z}^d$ and $k \in \mathbf{Z}$, we denote by $i + k$ the vertex $i + (k, 0, \cdots, 0)$. Further, with abuse of notation, if $0 \leq k \in \mathbf{Z}$ and $i \in \mathbf{Z}^d$, we define $[i, i+k] = \{i+l : 0 \le l \le k\}$, and we say that i is smaller than j if there exists an integer $k \ge 0$ with $j = i+k$. By our choice of M, $E[s(0)-n(0)] = L$ and so by the ergodic theorem, we have that for all $i \in \mathbf{Z}^d$,

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=i}^{i+k-1} (s(j) - n(j)) \to L \text{ a.s.} \tag{3}$$

The idea is to continue the proof as follows. Intuitively, we should be able to carry out the generation procedure of σ^* by using only the $W_{i,t}$, $i \in \mathbf{Z}^d$, $-M \le t \le -1$ because $E[s(j)] > E[n(j)]$ and so on the average, there are sufficiently many $W_{i,t}$'s around with $t \in [-M, -1]$; if, for some i, we need a $W_{i,t}$ with $t < -M$, we can transport unused $W_{j,t'}$'s from elsewhere with $t' \in [-M, -1]$. In this way, if this procedure is defined carefully, in a shift-invariant finitary manner, we should, by combining this with the procedure for generating σ^* above, obtain a finitary coding from the process $(W_i, i \in \mathbf{Z}^d)$ to μ, where $W_i = (W_{i,t}, -M \le t \le -1)$ (which is a finite-valued process). The above procedure will be carried out in stages. We now make the above intuition precise.

Let $\hat{W}_{j,t}, j \in \mathbf{Z}^d, t \in \{-M, \cdots, -1\}$ be i.i.d. random variables with the same distribution as $W_{O,0}$. We will construct a finitary coding from $\hat{W} = (\hat{W}_j,\, j \in \mathbf{Z}^d)$ to μ, where $\hat{W}_j = (\hat{W}_{j,t},\, t \in \{-M, \ldots, -1\})$. As suggested before, the method is to extend (if necessary) the $\hat{W}$-process to time indices less than $-M$ by using unneeded $\hat{W}_{i,t}$'s, $-M \le t \le -1$. First we define the notion 'being needed' and the variables $\hat{n}(i,t)$, $i \in \mathbf{Z}^d$, $t \in \{-(M+1), \cdots, -1\}$ exactly as before, but now with respect to the $\hat{W}$ process. Note that $\hat{n}(i,t)$ is measurable with respect to $\{\hat{W}_{j,s}, j \in \mathbf{Z}^d, s \in \{t+1, \ldots, -1\}\}$. Therefore, since the information that a certain (i,t) is needed tells us nothing about the value of $\hat{W}_{i,t}$, unneeded $\hat{W}_{i,t}$'s can be regarded as independent random variables with the 'correct' (original) distribution.

Let $T_1 = \mathbf{Z}^d \times (-\infty, -(M+1)]$ and $T_2 = \mathbf{Z}^d \times [-M, -1]$. For $p = (i,k) \in T_1 \cup T_2$, let

$$C_p = \{(i', k') : k+1 \le k' \le -1, \|i - i'\| \le 2|k|\}.$$

For $i \in \mathbf{Z}^d, \ell \ge 0$, let

$$A_{i\ell} = \{(j, r) \in T_1 : j \in [i, i+\ell]\}.$$

We will now define two processes $\{\hat{W}_p^n\}_{p \in T_1 \cup T_2, n \ge 0}$ and $\{\hat{S}_p^n\}_{p \in T_2, n \ge 0}$ on the same probability space.(These processes will in fact be defined in terms

of the variables $\{\hat{W}_j\}_{j\in\mathbf{Z}^d}$.) For $p\in T_2$, let $\hat{W}_p^n=\hat{W}_p$ for all $n\geq 0$. Informally (but not precisely), for $p\in T_1$, $\hat{W}_p^n$ will be
(a) "?" if at the end of stage n, we don't yet know if a $\hat{W}$ variable will be needed at space-time location p,
(b) "u" (for unneeded) if at the end of stage n, we know that a $\hat{W}$ variable will not be needed at space-time location p,
(c) "n" (for needed) if at the end of stage n, we know that a $\hat{W}$ variable will be needed at space-time location p but its value has not yet been determined, or
(d) some element of the alphabet A if at the end of stage n, it is known that a $\hat{W}$ variable will be needed at space-time location p and this variable is determined and given by $\hat{W}_p^n$.

Informally, for $p\in T_2$, $\hat{S}_p^n$ will be
(a) 1 if the space-time location p is not needed initially and if at the end of stage n, its value has not been "transported away" (and is therefore still available for use),
(b) 0 otherwise.

We now formally define these two processes inductively or in stages with respect to n. Each stage other than the 0th will consist of two substages.

STAGE 0

Let $\hat{W}_p^0=?$ for all $p\in T_1$ and $\hat{S}_p^0=I_{\{p \text{ is not needed }\}}$ for all $p\in T_2$.

STAGE 1(a)
Let $V^1=\{p\in T_1:\hat{W}_p^0=?\}$ (which is T_1 at this stage) and $U^1=\{p\in T_1:\hat{W}_p^0=\text{ "n" }\}$ (which is empty at this stage). We partition V^1 into three sets V_1^1,V_2^1,V_3^1 as follows. $V_1^1=\{p\in V^1:C_p\cap(V^1\cup U^1)\neq\emptyset\}$ (which is $\{(i,k)\in T_1:k<-(M+1)\}$ at this stage), $V_2^1=\{p\in V^1\backslash V_1^1:p\text{ is not needed }\}$ and $V_3^1=\{p\in V^1\backslash V_1^1:p\text{ is needed }\}$. (Note that if $p\in V^1\backslash V_1^1$, we can determine whether p is needed.) Let $\hat{W}^1$ be "?" on V_1^1, "u" on V_2^1 and "n" on V_3^1. Note that for each $i\in\mathbf{Z}^d$, there is at most one k such that $(i,k)\in V_3^1$.

STAGE 1(b)
We next partition $V_3^1\cup U^1(=V_3^1)$ into two sets V_4^1 and V_5^1 as follows. Let V_4^1 be

$$\{(i,k)\in V_3^1\cup U^1:\exists\ell\in[0,1]:\sum_{j\in[i,i+\ell],r\in[-M,-1]}\hat{S}_{j,r}^0\geq|(V_3^1\cup U^1)\cap A_{i\ell}|\}$$

and $V_5^1 = (V_3^1 \cup U^1)\backslash V_4^1$.

For $(i,k) \in V_4^1$, let ℓ be the minimal value satisfying the above and let q be the unique integer in $[-M,-1]$ such that

$$\sum_{j\in[i,i+\ell-1],r\in[-M,-1]} \hat{S}^0_{j,r} + \sum_{u=-M}^{q} \hat{S}^0_{i+\ell,u} = |(V_3^1 \cup U^1) \cap A_{i\ell}|.$$

Leave $\hat{W}^1$ unchanged on $V_1^1 \cup V_2^1 \cup V_5^1$ and for $(i,k) \in V_4^1$, let $\hat{W}^1_{i,k} = \hat{W}_{i+\ell,q}$ where ℓ and q are as above. (We think of the unneeded variable $\hat{W}_{i+\ell,q}$ being transported to space-time location (i,k).) Note that in light of the earlier remark that for each $i \in \mathbf{Z}^d$, there is at most one k such that $(i,k) \in V_3^1$, the random variable $\hat{W}_{i+\ell,q}$ is "transported" to at most one space-time point. If $p = (i+\ell,q) \in T_2$ corresponds to some $(i,k) \in V_4^1$ as above, then let $\hat{S}^1_p = 0$ and let $\hat{S}^1_p = \hat{S}^0_p$ otherwise.

STAGE 2(a)

Let $V^2 = \{p \in T_1 : \hat{W}^1_p =?\}$ and $U^2 = \{p \in T_1 : \hat{W}^1_p =$ "n" $\}$. On $T_1\backslash(V^2 \cup U^2)$, let $\hat{W}^2 = \hat{W}^1$. We partition V^2 into three sets V_1^2, V_2^2, V_3^2 as follows. $V_1^2 = \{p \in V^2 : C_p \cap (V^2 \cup U^2) \neq \emptyset\}$, $V_2^2 = \{p \in V^2\backslash V_1^2 : p$ is not needed $\}$ and $V_3^2 = \{p \in V^2\backslash V_1^2 : p$ is needed $\}$. (As in stage 1(a), note that, if $p \in V^2\backslash V_1^2$, we can determine from previous information whether p is needed, where 'needed' is defined in the analogous way as before). Let $\hat{W}^2$ be "?" on V_1^2, "u" on V_2^2 and "n" on V_3^2.

STAGE 2(b)

We next partition $V_3^2 \cup U^2$ into two sets V_4^2 and V_5^2 as follows. Let V_4^2 be

$$\{(i,k) \in V_3^2 \cup U^2 : \exists \ell \in [0,2] : \sum_{j\in[i,i+\ell],r\in[-M,-1]} \hat{S}^1_{j,r} \geq |(V_3^2 \cup U^2) \cap A_{i\ell}|\}$$

and $V_5^2 = (V_3^2 \cup U^2)\backslash V_4^2$.

For $(i,k) \in V_4^2$, let ℓ be the minimal value satisfying the above and let q be the unique integer in $[-M,-1]$ such that

$$\sum_{j\in[i,i+\ell-1],r\in[-M,-1]} \hat{S}^1_{j,r} + \sum_{u=-M}^{q} \hat{S}^1_{i+\ell,u} = |(V_3^2 \cup U^2) \cap A_{i\ell}|.$$

Leave $\hat{W}^2$ unchanged on $V_1^2 \cup V_2^2 \cup V_5^2$ and for $(i,k) \in V_4^2$, let $\hat{W}^2_{i,k} = \hat{W}_{i+\ell,q}$ where ℓ and q are as above. Note again that the random variable $\hat{W}_{i+\ell,q}$ is "transported" to at most one space-time point. If $p = (i+\ell,q) \in T_2$

corresponds to some $(i,k) \in V_4^2$ as above, then let $\hat{S}_p^2 = 0$ and let $\hat{S}_p^2 = \hat{S}_p^1$ otherwise.

STAGE n(a)
Let $V^n = \{p \in T_1 : \hat{W}_p^{n-1} = ?\}$ and $U^n = \{p \in T_1 : \hat{W}_p^{n-1} =$ "n" $\}$. On $T_1 \backslash (V^n \cup U^n)$, let $\hat{W}^n = \hat{W}^{n-1}$. We partition V^n into three sets V_1^n, V_2^n, V_3^n as follows. $V_1^n = \{p \in V^n : C_p \cap (V^n \cup U^n) \neq \emptyset\}$, $V_2^n = \{p \in V^n \backslash V_1^n :$ p is not needed $\}$ and $V_3^n = \{p \in V^n \backslash V_1^n : p$ is needed $\}$. Let $\hat{W}^n$ be "?" on V_1^n, "u" on V_2^n and "n" on V_3^n.

STAGE n(b)
We partition $V_3^n \cup U^n$ into two sets V_4^n and V_5^n as follows. Let V_4^n be

$$\{(i,k) \in V_3^n \cup U^n : \exists \ell \in [0,n] : \sum_{j \in [i,i+\ell], r \in [-M,-1]} \hat{S}_{j,r}^{n-1} \geq |(V_3^n \cup U^n) \cap A_{i\ell}|\}$$

and $V_5^n = (V_3^n \cup U^n) \backslash V_4^n$.

For $(i,k) \in V_4^n$, let ℓ be the minimal value satisfying the above and let q be the unique integer in $[-M,-1]$ such that

$$\sum_{j \in [i,i+\ell-1], r \in [-M,-1]} \hat{S}_{j,r}^{n-1} + \sum_{u=-M}^{q} \hat{S}_{i+\ell,u}^{n-1} = |(V_3^n \cup U^n) \cap A_{i\ell}|.$$

Leave $\hat{W}^n$ unchanged on $V_1^n \cup V_2^n \cup V_5^n$ and for $(i,k) \in V_4^n$, let $\hat{W}_{i,k}^n = \hat{W}_{i+\ell,q}$ where ℓ and q are as above. Note that as before, the random variable $\hat{W}_{i+\ell,q}$ is "transported" to at most one space-time point. If $p = (i+\ell, q) \in T_2$ corresponds to some $(i,k) \in V_4^n$ as above, then let $\hat{S}_p^n = 0$ and let $\hat{S}_p^n = \hat{S}_p^{n-1}$ otherwise. This completes the construction of the two processes $\{\hat{W}_p^n\}_{p \in T_1 \cup T_2, n \geq 0}$ and $\{\hat{S}_p^n\}_{p \in T_2, n \geq 0}$.

The idea is now to use the $\{\hat{W}_p^n\}_{p \in T_1 \cup T_2, n \geq 0}$ variables to construct a $\hat{\sigma}^*$ analogously to what we did earlier. In doing this, we need to know that for all $p \in T_1$, $\hat{W}_p^n$ is either "u" or some value in A for sufficiently large n. To do all this, it is useful to construct analogous processes to the above but with respect to the original variables $\{W_p\}_{p \in T_1 \cup T_2}$. This will allow a more precise comparison.

We now define two processes $\{W_p^n\}_{p \in T_1 \cup T_2, n \geq 0}$ and $\{S_p^n\}_{p \in T_2, n \geq 0}$ which are measurable with respect to $\{W_p\}_{p \in T_1 \cup T_2}$ and which are defined almost completely analogously to $\{\hat{W}_p^n\}_{p \in T_1 \cup T_2, n \geq 0}$ and $\{\hat{S}_p^n\}_{p \in T_2, n \geq 0}$ but with one essential difference. The processes $\{W_p^n\}_{p \in T_1 \cup T_2, n \geq 0}$ and $\{S_p^n\}_{p \in T_2, n \geq 0}$ are defined inductively over n exactly as $\{\hat{W}_p^n\}_{p\in T_1\cup T_2, n\geq 0}$ and $\{\hat{S}_p^n\}_{p\in T_2, n\geq 0}$ except that when we are at some stage $n(b)$ where we are about to assign a value from A to a space-time location $p \in T_1$, rather than transporting the value from a space-time location in T_2, we let W_p^n simply be W_p; i.e., we reveal the value which was already there (but which we "did not know" before this stage). Since transporting a variable tells us nothing about its value, it is clear that

$$\{\hat{W}_p^n, \hat{S}_{p'}^n\}_{p\in T_1\cup T_2, p'\in T_2, n\geq 0} =_{\mathcal{D}} \{W_p^n, S_{p'}^n\}_{p\in T_1\cup T_2, p'\in T_2, n\geq 0} \tag{4}$$

where $=_{\mathcal{D}}$ means equal in distribution.

Next, note that it follows from the construction that for a fixed $p \in T_1$, and for *every* ω, $\hat{W}_p^n$ as a function of n must behave in one of the following ways.

(i) start off with value "?" and remain fixed forever.

(ii) start off with value "?", change to "u" at some point and then remain fixed forever.

(iii) start off with value "?", change to a value in A at some point and then remain fixed forever.

(iv) start off with value "?", change to "n" at some point and then remain fixed forever.

(v) start off with value "?", change to "n" at some point and then change to a value in A at some later point and then remain fixed forever.

Lemma 3.7 *For all $p \in T_1$, behavior (i) and (iv) above do not occur a.s.*

Proof of Lemma 3.7: First note that for *every* ω, if behavior (i) occurs for some $p \in T_1$, then behavior (iv) must occur for some p'. To see this, let $p = (i, k)$ be a point where $\hat{W}_p^n$ is always in state "?" and where k is maximal with respect to this property. Then $p \in V_1^n$ for all n and so there must be a $p' \in C_p$ where $\hat{W}_{p'}^n$ remains fixed in state "n". Hence we need only rule out behavior (iv) and we do this for a fixed $p_0 = (i_0, k_0) \in T_1$.

In view of (4), it suffices to do this for $\{W_{p_0}^n\}_{n\geq 0}$ instead. First, choose n_0 such that $W_{p_0}^{n_0} =$ "n" from which it immediately follows that $p_0 \in V_3^{n_0} \cup U^{n_0}$. Now by (3), choose k_0 such that

$$\sum_{j=i_0}^{i_0+k_0} (s(j) - n(j)) > n_0 M,$$

and choose $n_0' > \max\{n_0, k_0\}$. (Note that $n_0 M$ is the maximum number of space-time points p in $[i_0, i_0+\infty) \times [-M, -1]$ which have been "transported" to a point to the left of i_0 by stage n_0.) We claim that $W_{p_0}^{n'_0} \in A$, which would complete the proof. It is clear from the construction that no point in T_2 at or to the right of p_0 can after time n_0 be "transported" to something to the left of p_0 as long as p_0 remains in state "n". Therefore if at the end of stage $n'_0 - 1$, p_0 is still in state "n", then it easily follows from the fact that

$$\sum_{j=i_0}^{i_0+k_0} s(j) > n_0 M + \sum_{j=i_0}^{i_0+k_0} n(j)$$

that

$$\sum_{j\in[i_0,i_0+k_0],r\in[-M,-1]} S_{j,r}^{n'_0-1} \geq |(V_3^{n'_0} \cup U^{n'_0}) \cap A_{i_0 k_0}|,$$

which implies that $W_{p_0}^{n'_0} \in A$, as desired. $\Box$

Letting

$$\hat{W}_p^\infty := \lim_{n\to\infty} \hat{W}_p^n \text{ and } W_p^\infty := \lim_{n\to\infty} W_p^n$$

(which are clearly well-defined since all the sequences are eventually constant), Lemma 3.7 implies that a.s. $\hat{W}_p^\infty$ is "u" or an element of A for all $p \in T_1 \cup T_2$ while (4) immediately yields

$$\{\hat{W}_p^\infty\}_{p\in T_1} =_{\mathcal{D}} \{W_p^\infty\}_{p\in T_1}. \tag{5}$$

It is clear from the construction that W_p^∞ is W_p if p is needed and "u" otherwise. Because of this and (5), it immediately follows that if we define $\hat{\sigma}^*$ as we defined σ^* before but now with respect to the $\hat{W}_p^\infty$ variables, then $\hat{\sigma}^*$ will have the correct distribution μ. (Note that the $\hat{W}_p^\infty$ variables are not always in A, since some of them take the value "u" but this obviously does not matter since such a variable is unneeded in this case).

The composition of first going from the $\{\hat{W}_j\}_{j\in\mathbf{Z}^d}$ random variables to the $\{\hat{W}_p\}_{p\in T_1\cup T_2}$ random variables and then to $\hat{\sigma}^*$ yields a stationary coding from $\{\hat{W}_j\}_{j\in\mathbf{Z}^d}$ to μ and now we need only show that it is finitary. To do this, it clearly suffices to show that for all $\epsilon > 0$, there exists N such that the probability that $\hat{\sigma}_0^*$ is determined by $\{\hat{W}_j\}_{\|j\|\leq N}$ is $> 1-\epsilon$. For this, first choose N_1 such that $P(\tau_0 > N_1) < \epsilon/2$. Let $S_{N_1} = \{(i,t) : -N_1 \leq t \leq -1, \|i\| \leq |t|\}$. Hence σ_0^* is determined by the random variables $\{W_p\}_{p\in S_{N_1}}$ with probability $> 1-\epsilon/2$ and so by the above, $\hat{\sigma}_0^*$ is determined by the random variables $\{\hat{W}_p^\infty\}_{p\in S_{N_1}}$ with probability $> 1-\epsilon/2$. Next, choose N_2 such that with probability $> 1-\epsilon/2$, $\hat{W}_p^\infty = \hat{W}_p^{N_2}$ for all $p \in S_{N_1}$. This would imply that the $\hat{W}_p^\infty$ variables for $p \in S_{N_1}$ are determined by stage N_2 with probability $> 1 - \epsilon/2$. Let $N = N_1 + N_2$. Clearly, by construction, $\{\hat{W}_p^{N_2}\}_{p \in S_{N_1}}$ is measurable with respect to $\{\hat{W}_j\}_{\|j\| \leq N}$. Therefore the random variables $\{\hat{W}_j\}_{\|j\| \leq N}$ determine $\hat{\sigma}_0^*$ with probability $> 1 - \epsilon$, as desired. $\square$

4 Application to the Ising model

We consider the ferromagnetic Ising model on $\mathbf{Z}^d$, $d \geq 2$, with interaction parameter $J, J > 0$, and zero external field, that is, we consider random fields $\{\sigma_j\}_{j \in \mathbf{Z}^d}$ taking values ± 1 with the property that, for each $i \in \mathbf{Z}^d$, and $\alpha \in \{-1, +1\}$,

$$P(\sigma_i = \alpha \mid \sigma_j, j \neq i) = p_i^{(\sigma_j, j \sim i)}(\alpha), \tag{6}$$

where we have used the following notation: $j \sim i$ means that i and j are neighbors, and

$$p_i^\gamma(\alpha) = \frac{\exp(\alpha J \sum_{j \sim i} \gamma_j)}{\exp(\alpha J \sum_{j \sim i} \gamma_j) + \exp(-\alpha J \sum_{j \sim i} \gamma_j)}, \gamma \in \{-1, +1\}^{\{j : j \sim i\}}. \tag{7}$$

We call such distributions Ising distributions (with parameter J).

It is well-known (see [25], p. 189–190) that there exists a critical value $J_c(d) \in (0, \infty)$ such that if $J < J_c(d)$ then there is a unique Ising distribution in d dimensions, while for $J > J_c(d)$ there is more than one Ising distribution in d dimensions. This corresponds with the occurrence of a so-called *phase transition*. In particular, if we assign the value $+1$, respectively -1, to each vertex outside $B_n = [-n, n]^d \cap \mathbf{Z}^d$, and consider the (unique) distribution which satisfies (6) for each i inside the cube, then, by letting $n \to \infty$ and taking weak limits, we obtain two Ising distributions, called respectively the 'plus state' and the 'minus state' of the Ising model, which are equal when $J < J_c(d)$ but distinct when $J > J_c(d)$. The existence of the above limits follows from well-known stochastic monotonicity results (see [25], p.189). We now restate Theorem 1.1.

Theorem 4.1 *Consider the Ising model defined by (6).*
a) If $J < J_c(d)$, the plus state (which equals the minus state) of the Ising model is a finitary factor of a finite-valued i.i.d. process.
b) If $J > J_c(d)$, the plus state (and the minus state) of this model is not a finitary factor of a finite-valued i.i.d. process.

PROOF:

b) This follows immediately from Theorem 2.1.

a) It is well-known, and can be easily proved by (now standard) monotonicity and coupling arguments, that, if $J < J_c(d)$, then under the following continuous-time dynamics, the distribution at time t starting from any configuration converges to the (unique) Ising distribution: Each vertex has a clock which rings after i.i.d. exponentially distributed (parameter 1) time intervals. All these clocks behave independently of each other. When a clock rings (say, at vertex i, at time t), then the spin value of i is updated, i.e., is replaced by a new value, which is drawn (independently of anything else) from the distribution $p_i^{(\sigma_j(t), j \sim i)}$. Here $\sigma_j(t)$ denotes the spin value at j at time t. A much deeper result, proved by Martinelli and Olivieri in [28], which, in turn, involves a key result in [2] about the spatial mixing properties of the Ising distribution (see also [15]) is that the above mentioned convergence occurs exponentially. This is of crucial importance for us. Instead of formulating the Martinelli-Olivieri result precisely here, we will state its analog for the following, *discrete-time* dynamics, which corresponds with a cellular automaton to which Theorem 3.4 can be applied. (There are several other ways to set up a discrete-time dynamics for the Ising model (see e.g. [40] and [27]) but for our purpose we prefer the one below). At each discrete time step, each vertex is, independent of the others, 'activated' with probability 1/2. When a vertex is activated, but none of its neighbors is, then its value is updated (where 'updating' means the same as in the continuous case). Again one can easily verify (by the same standard arguments referred to above) that if $J < J_c(d)$, the distribution starting from any configuration converges to the unique Ising distribution as $t \to \infty$. This discrete dynamics can be described in a semi-deterministic way (like the PCA in Section 3), which has the advantage that it couples the time evolutions for all initial configurations. Such a coupling is analogous to the so-called 'basic coupling' (see [25], p.124). Let $A_{i,t}, i \in \mathbf{Z}^d$, $t \in \mathbf{N}$, be i.i.d. with $P(A_{i,t} = 1) = 1 - P(A_{i,t} = 0) = 1/2$. As before, let $\partial i = \{j : j \sim i\}$. Let $U'_{i,t}, i \in \mathbf{Z}^d, t \in \mathbf{N}$ be i.i.d. uniformly distributed random variable on the interval $(0, 1)$ and independent of the above A-process. Next, define $U_{i,t} = \max\{p_O^\gamma(1) : \gamma \in \{-1,+1\}^{\partial O}, \, p_O^\gamma(1) < U'_{i,t}\}$ (the maximum here is interpreted as 0 if $p_O^\gamma(1) \geq U'_{i,t}$ for all $\gamma \in \{-1,+1\}^{\partial O}$). Finally, define $W_{i,t} = (A_{i,t}, U_{i,t})$.

Consider the following PCA with initial configuration ω.

$$
\begin{aligned}
\sigma(\omega,0) &= \omega. && (8)\\
\sigma_i(\omega,t+1) &= \sigma_i(\omega,t) \text{ if } A_{i,t}=0 \text{ or } A_{j,t}=1 \text{ for some } j\sim i;\\
&= +1\ (-1) \text{ if } U_{i,t} <\ (\geq)\ p_i^{(\sigma_j(\omega,t),j\sim i)}(1) \text{ otherwise.}
\end{aligned}
$$

One can easily verify that this PCA is monotone and corresponds with the discrete dynamics described above. Next, the following proposition is the discrete-time analog of Theorem 5.1 in [28]. The latter uses Theorem 3.1 in the same paper, of which the statement and proof go essentially (and straightforwardly) through step-by-step for the discrete-time dynamics described above.

Proposition 4.2 *If $J < J_c(d)$, then the above PCA is exponentially ergodic.*

Theorem 4.1 a) now follows immediately from Theorem 3.4, Proposition 4.2 and the earlier observed facts that this PCA is monotone and has the Ising distribution as its limit distribution. □

Remark(k): The Ising model is just one (important) example to illustrate Theorem 3.4. It should be clear from this example that the combination of our Theorem 3.4 with the before mentioned Theorem 3.1 in [28] yields analogs of Theorem 4.1 (a) for a large class of random fields.

Remark(l): It has been proved (see [3] and p. 171-172 in [11]) that for all dimensions except 3, (and is also certainly believed in 3 dimensions) for $J = J_c(d)$ the plus state and the minus state of the Ising model are equal. However, it is well-known that at the critical point the dynamics cannot be exponentially ergodic, so that the question whether there exists a finitary coding for the critical Ising model is left open. However, one can prove the following result which was obtained jointly with Yuval Peres.

For a finitary mapping, for $x \in \mathbf{Z}^d$, we let N_x denote the random variable which is the side length of the smallest hypercube about the point x which has the property that the value at x in the process we are mapping to is determined by the values of the configuration that we are mapping from in this hypercube. (Finitary means that N_x is finite a.s.) N_0 is called the *coding length* in one dimension.

Theorem 4.3 *There does not exist a finitary factor map from a finite-valued i.i.d. process to the plus state for the Ising model at the critical value in d dimensions which has finite expected coding volume, i.e. for which $E[N_0^d] < \infty$.*

Proof of Theorem 4.3: Let $\{\sigma_x\}_{x\in\mathbf{Z}^d}$ denote the Gibbs state at the critical value in d dimensions. We may assume that it is unique, since otherwise there is no finitary mapping at all to the plus state by Theorem 2.1. It is well known (and goes back to [38] and [24]) that in this case $\sum_{j\in\mathbf{Z}^d} E[\sigma_0\sigma_j]=\infty$. We will show that if there is a finitary factor map with finite expected coding volume, then the above sum is finite, which completes the proof. (Recall again that $E[\sigma_x\sigma_y]\geq 0$ for all $x,y\in\mathbf{Z}^d$ and so we don't need to take absolute values here.)

Fix $j\in\mathbf{Z}^d$. Then $E[\sigma_0\sigma_j]$ is equal to

$$
\begin{aligned}
&\sum_{\substack{k,\ell\geq 0\\ \max\{k,\ell\}\geq\lfloor\frac{|j|}{2}\rfloor-1}} E[\sigma_0 I_{\{N_0=k\}}\sigma_j I_{\{N_j=\ell\}}] \quad + \\
&\sum_{\substack{k,\ell\geq 0\\ \max\{k,\ell\}<\lfloor\frac{|j|}{2}\rfloor-1}} E[\sigma_0 I_{\{N_0=k\}}\sigma_j I_{\{N_j=\ell\}}] \quad \leq \\
&\sum_{\substack{k,\ell\geq 0\\ \max\{k,\ell\}\geq\lfloor\frac{|j|}{2}\rfloor-1}} P(\{N_0=k\}\cap\{N_j=\ell\}) \quad + \\
&\sum_{\substack{k,\ell\geq 0\\ \max\{k,\ell\}<\lfloor\frac{|j|}{2}\rfloor-1}} E[\sigma_0 I_{\{N_0=k\}}]E[\sigma_j I_{\{N_j=\ell\}}] \quad \leq \\
&2\sum_{k\geq\lfloor\frac{|j|}{2}\rfloor-1} P(\{N_0=k\}) \quad + \\
&\left(\sum_{k=0}^{\lfloor\frac{|j|}{2}\rfloor-2} E[\sigma_0 I_{\{N_0=k\}}]\right)^2 \quad = \\
&2\sum_{k\geq\lfloor\frac{|j|}{2}\rfloor-1} P(\{N_0=k\}) \quad + \\
&\left(\sum_{k\geq\lfloor\frac{|j|}{2}\rfloor-1} E[\sigma_0 I_{\{N_0=k\}}]\right)^2 \quad \leq \\
&3\sum_{k\geq\lfloor\frac{|j|}{2}\rfloor-1} P(\{N_0=k\}).
\end{aligned}
$$

The equality above was based on the fact that $E[\sigma_0]=0$. Since it is clear that

$$\sum_{j\in\mathbf{Z}^d}\sum_{k\geq\lfloor\frac{|j|}{2}\rfloor-1}P(\{N_0=k\})$$

is bounded above by a constant times $E[(N_0)^d]$, this completes the proof. $\Box$

This leaves open two questions.

Question 1: Does there exist a finitary factor map from a finite-valued i.i.d. process to the plus state for the Ising model at the critical value in d dimensions?

Question 2: Does the finitary coding we construct for the subcritical Ising Model have finite expected coding volume and if not, does there exist such a finitary coding?

Note that if the answer to both questions is yes, then the behavior (with respect to finitary coding) at the critical value will be different from both the subcritical and supercritical cases.

Although we feel that when studying finitary codings, one should primarily be interested in the case where the i.i.d. process that one is mapping from is finite-valued, we mention that in the larger category where the i.i.d. process is not necessarily finite-valued, then we can already distinguish the critical Ising model for $d \neq 3$ from both the subcritical and supercritical cases. This is because of the following. It is not hard to show (using similar ideas as earlier) that the Ising model above the critical value is not a finitary factor of any i.i.d. process (finite-valued or otherwise). However in the critical case, for $d \neq 3$, there is such a mapping by combining Lemmas 3.5, 3.6 and Remarks (i) and (l), showing that the critical case is different from the supercritical case. Next, it is clear that the finitary code given in Remark (i) (where the domain is an uncountably valued i.i.d. process) has finite expected coding volume in the subcritical case. On the other hand, the proof of Theorem 4.3 immediately extends to the case where the i.i.d. process is not finite-valued showing that the critical case is also different from the subcritical case.

5 Some further open questions

Here we mention a few open questions. In view of Theorem 2.1, one may ask whether ν being the unique translation invariant Markov random field with its conditional probabilities implies that ν is a finitary factor of an i.i.d. process. This is not true as we will indicate below. As noted in the proof of the theorem, such a finitary factor satisfies the blowing-up property. In [30], it is proved that in 1 dimension a process satisfies the blowing–up property if and only if it is a Bernoulli shift, has the "exponential rate of convergence property for frequencies" (which means that the mean ergodic theorem for cylinder sets occurs at an exponential rate) and has the "exponential rate of convergence property for entropy". This proof extends to $d \geq 2$ dimensions. So there are potentially three natural ways to look for counterexamples, and we discuss two of them.

We first mention that there exists a translation invariant Markov random field which is the unique translation invariant Markov random field with its conditional probabilities, but which is *not* a Bernoulli shift. For example, for the antiferromagnetic Ising Model in 2 dimensions (see [14] for the definition of this) above the critical interaction parameter, there is a unique translation invariant Markov random field with the proper conditional probabilities, but this field is not even totally ergodic and hence not a Bernoulli shift (it splits as a convex combination of two periodic Markov random fields with the proper conditional probabilities). This gives our first counter-example, as mentioned above. What is happening here is that there may be *nontranslation invariant* Markov random fields (even nonperiodic Markov random fields can exist, e.g., the Ising model in 3 dimensions (see [10])) and one needs to know whether the unique translation invariant Markov random field that one is looking at is extremal within the class of *all* Markov random fields with the given conditional probabilities, not just within the class of the translation invariant ones. Further, if we do assume that in addition to ν being the unique translation invariant Markov random field with its conditional probabilities, it is also extremal in the class of all Markov random fields with its conditional probabilities, then it is not known whether ν is necessarily a Bernoulli shift (although it is in this case necessarily K, see [14]). In fact, even if ν is the unique Markov random field with its conditional probabilities (unique among all both translation invariant and nontranslation invariant Markov random fields), which implies that ν is translation invariant and K, it is still not known whether ν is necessarily a Bernoulli shift. Here it is very important to point out that the example given in [16] has the property that it is *not* the unique Markov random field with its conditional probabilities.

Next, as far as the second property of "exponential rate of convergence property for frequencies", a measure ν which is the unique translation invariant Markov random field with strictly positive conditional probabilities does in fact satisfy this property. This follows from Theorem 4.1 in [12], the lower semicontinuity of relative entropy and Theorem 15.37 in [14], p. 323.

The above discussion leads to the following question.

Question 3: If a translation invariant Markov random field μ is the unique Markov random field (among both translation invariant and non-translation invariant fields) with its conditional probabilities, is μ necessarily a Bernoulli shift? (It is K by the above discussion).

Acknowledgement. We thank Ronald Meester for comments on an earlier version of this paper, Yuval Peres for jointly proving Theorem 4.3 with us, and a referee for some suggestions. Part of this research has been carried out when the authors participated, with financial support from Rutgers University, in the DIMACS program on Discrete Probability in the spring of 1997. J. van den Berg also thanks the Stochastic Centre at Chalmers University of Technology for financial support to participate in the program on Percolation, Particle Systems and Ergodic Theory in the fall of 1997. J.E. Steif was partly supported by grants from the Swedish Natural Science Research Council and from the Royal Swedish Academy of Sciences.

References

[1] Adams, S., (1992). Fölner Independence and the amenable Ising model, *Ergodic Theory Dynam. Systems*, **12**, 633-657.

[2] Aizenman, M., Barsky, D. J. and Fernandez, R., (1987). The phase transition in a general class of Ising-type models is sharp, *J. Statist Phys.*, **47**, 343–374.

[3] Aizenman, M., and Fernandez, R., (1986). On the critical behavior of the magnetization in high–dimensional Ising models, *J. Statist Phys.*, **44**, 393–454.

[4] Akcoglu, M. A., del Junco, A. and Rahe, M., (1979). Finitary codes between Markov processes, *Z. Wahrsch. Verw. Gebiete*, **47**, 305–314.

[5] Blum, J. R. and Hanson, D., (1963). On the isomorphism problem for Bernoulli schemes, *Bull. Amer. Math. Soc.*, **69**, 221–223.

[6] Burton, R. M. and Steif, J. E., (1994). Nonuniqueness of measures of maximal entropy for subshifts of finite type, *Ergodic Theory Dynam. Systems*, **14**, 213–235.

[7] Burton, R. M. and Steif, J. E., (1995). New results on measures of maximal entropy, *Isr. J. Math.*, **89**, 275–300.

[8] Burton, R. M. and Steif, J. E., (1996). Some 2-d symbolic dynamical systems: entropy and mixing, *Ergodic Theory and Dynamical Systems (with applications to* $\mathbf{Z}^n$*-actions, number theory, statistical mechanics, and algebra)*, eds., Pollicott M. and Schmidt, K, 297–305.

[9] Csiszar, I. and Korner, J., (1981). *Information theory: Coding theorems for discrete memoryless systems,* Academic Press, Inc., New York.

[10] Dobrushin, R. L., (1972). Gibbs states describing coexistence of phases for a three–dimensional Ising model, *Th. Prob. Appl.*, **17**, 582–600.

[11] Ellis, R. S., (1985) *Entropy, Large Deviations, and Statistical Mechanics*, Springer–Verlag, New York.

[12] Föllmer, H. and Orey, S., (1988). Large deviations for the empirical field of a Gibbs measure, *Ann. Probab.* **16**, 961–977.

[13] Friedman, N. A. and Ornstein, D., (1971). On Isomorphism of Weak Bernoulli Transformations, *Adv. Math.*, **5**, 365-394.

[14] Georgii, H., (1988). *Gibbs measures and phase transitions,* de Gruyter, New York.

[15] Higuchi, Y., (1993). Coexistence of infinite ($*$)-clusters. II. Ising percolation in two dimensions, *Probab. Th. Rel. Fields* **97**, 1–33.

[16] Hoffman, C., (1998). A Markov random field which is K but not Bernoulli, Israel J. Math. to appear.

[17] den Hollander, F. and Steif, J. E., (1997). On K–automorphisms, Bernoulli shifts and Markov random fields, *Ergodic Theory Dynam. Systems*, **17**, Part 2, 405–415.

[18] den Hollander, F. and Steif, J. E., (1998). On the equivalence of certain ergodic properties for Gibbs states, *Ergodic Theory Dynam. Systems*, to appear.

[19] del Junco, A., (1980). Finitary coding of Markov random fields, *Z. Wahrsch. Verw. Gebiete*, **52**, 193–202.

[20] Katznelson, Y. and Weiss, B., (1972). Commuting measure-preserving transformations, *Isr. J. Math.*, **12** 161–173.

[21] Keane, M. and Smorodinsky, M., (1977). A class of finitary codes, *Isr. J. Math.*, **26**, 352–371.

[22] Keane, M. and Smorodinsky, M., (1979). Bernoulli schemes of the same entropy are finitarily isomorphic, *Ann. Math.*, **109**, 397–406.

[23] Keane, M. and Smorodinsky, M., (1979). Finitary isomorphisms of irreducible Markov shifts, *Isr. J. Math.*, **34**, 281–286.

[24] Lieb, E.H., (1980). A refinement of Simon's correlation inequality, *Comm. Math. Phys.*, **77**, 127–135.

[25] Liggett, T. M., (1985). *Interacting Particle Systems*, Springer-Verlag, Berlin.

[26] Lind, (1982). Ergodic group automorphisms are exponentially recurrent, *Isr. J. Math.*, **41**, 313–320.

[27] Maes, C. and Shlosman, S.B. (1993). When is an interacting particle system ergodic? *Comm. Math. Phys.*, **151**, 447–466.

[28] Martinelli, F. and Olivieri, E., (1994). Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.*, **161**, 447–486.

[29] Marton, K., The blowing–up property for processes with memory, submitted.

[30] Marton, K. and Shields, P., (1994). The positive-divergence and blowing–up properties, *Isr. J. Math.*, **86**, 331–348.

[31] Meshalkin, L. D., (1959). A Case of Isomorphism of Bernoulli Schemes, *Dokl. Akad. Nauk. SSSR*, **128**, 41–44.

[32] Monroy, G. and Russo, B., (1975). A family of codes between some Markov and Bernoulli schemes, *Comm. Math. Phys.*, **43**, 155–159.

[33] Ornstein, D.S., (1974). *Ergodic Theory, Randomness and Dynamical Systems*, Yale University Press New Haven.

[34] Ornstein, D.S. and Weiss, B., (1973). $\mathbf{Z}^d$-actions and the Ising model, unpublished manuscript.

[35] Ornstein, D. S. and Weiss, B., (1987). Entropy and isomorphism theorems for actions of amenable groups, *J. Anal. Math.*, **48**, 1–141.

[36] Propp, J.G. and Wilson, D.B., (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures Algorithms*, **9**, 223–252.

[37] Schmidt, K., (1995). *Dynamical Systems of Algebraic Origin*, Birkhäuser.

[38] Simon, B., (1980). Correlation inequalities and the decay of correlations in ferromagnets, *Comm. Math. Phys.*, **77**, 111–126.

[39] Slawny, J., (1981). Ergodic Properties of Equilibrium States, *Comm. Math. Phys.*, **80**, 477–483.

[40] Steif, J.E., (1988). The ergodic structure of interacting particle systems. Stanford University Ph.D. Thesis.

[41] Talagrand, M., (1995). Concentration of measure and isoperimetric inequalities in product spaces, *Publications I.H.E.S.*, **81**, 73–205.