The Annals of Probability
1991, Vol. 19, No. 2, 609–635

SPACE-TIME BERNOULLICITY OF THE LOWER AND UPPER STATIONARY PROCESSES FOR ATTRACTIVE SPIN SYSTEMS[1]

BY JEFFREY E. STEIF

Cornell University

In this paper, we study spin systems, probabilistic cellular automata and interacting particle systems, which are Markov processes with state space $\{0,1\}^{\mathbf{Z}^n}$. Restricting ourselves to attractive systems, we consider the stationary processes obtained when either of two distinguished stationary distributions is used, the smallest and largest stationary distributions with respect to a natural partial order on measures. In discrete time, we show that these stationary processes with state space $\{0,1\}^{\mathbf{Z}^n}$ and index set $\mathbf{Z}$ are isomorphic (in the sense of ergodic theory) to an independent process indexed by $\mathbf{Z}$. In the translation invariant case, we prove the stronger fact that these stationary processes, viewed as $\{0,1\}$-valued processes with index set $\mathbf{Z}^n \times \mathbf{Z}$ (space-time), are isomorphic to an independent process also indexed by $\mathbf{Z}^n \times \mathbf{Z}$. Such processes are called Bernoulli shifts. Finally, we extend all of these results to continuous time.

1. Introduction. In this paper, we consider both discrete time spin systems or probabilistic cellular automata (PCA) and continuous time spin systems or interacting particle systems (IPS). These will be Markov processes with state space $X = \{0,1\}^{\mathbf{Z}^n}$, the set of configurations of 0's and 1's on the n-dimensional lattice. The transitions will be governed by a family of functions

$$\{c(x,\eta)\}_{x \in \mathbf{Z}^n,\, \eta \in X},$$

where

$$B = \sup_{x \in \mathbf{Z}^n,\, \eta \in X} c(x,\eta) \le \tfrac{1}{2}$$

and $c(x,\eta)$ is, for fixed x, a continuous function of η where X is given the product topology. The reason for having the $\frac{1}{2}$ bound rather than the more natural bound of 1 is explained later.

These two conditions plus one more which is given below are the three conditions in Liggett (1985) imposed on the spin rates of a spin system to insure the existence of a continuous time process. This third condition is

$$\sup_{x \in \mathbf{Z}^n} \sum_{y \neq x} \sup_{\eta \in X} |c(x,\eta) - c(x,\eta^y)| < \infty,$$

where η^y is the same configuration as η except switched at lattice point y. We shall call this last quantity M, which we assume to be finite. The summands in

Received May 1989; revised March 1990.

[1]Research supported by Stanford University, the Alfred P. Sloan Foundation and an NSF Mathematical Sciences Postdoctoral Research Fellowship.

AMS 1980 *subject classifications*. Primary 28D15, 60G10, 60K35.

Key words and phrases. Bernoullicity, attractive spin systems, couplings, $\bar{d}$-metric.

the above can be interpreted as the effect of lattice point y on the spin rate at lattice point x. Actually in Liggett (1985), any uniform bound on the functions $\{c(x, \eta)\}$ together with the previous assumptions suffices for the existence of the continuous time process.

In Sections 2–4 of this paper, we deal exclusively with discrete time and do not discuss the continuous time situation until Section 5 except for minor comments. To obtain a process in discrete time, we only need to make the first two previous assumptions. We make no general assumption of translation invariance although we make this assumption later.

The evolution of our process in discrete time is defined as follows. If the state of the system is η, then at the next stage each lattice point x switches its value independently with probability $c(x, \eta)$. In particular, this yields a product measure at the next stage, which we denote by $T\eta$. We can then evolve any initial distribution ν and we denote this evolved measure by $T\nu$ where of course, $T\nu = \int_X T\eta \, d\nu(\eta)$. For the definition of the continuous time process for the above spin rates, see Liggett (1985). Informally, the process when in state η waits an exponential amount of time with parameter $c(x, \eta)$ and then switches its value at x. In continuous time, we let $T^t\nu$ denote the distribution at time t when ν is the initial distribution. In discrete time, μ is called *stationary* if $T\mu = \mu$, while in continuous time, one requires that $T^t\mu = \mu$ for all t.

In this paper, we deal exclusively with attractive systems. We note that X has a natural partial order defined on it: $\eta \preccurlyeq \delta$ if $\eta(x) \le \delta(x)$ for all $x \in \mathbf{Z}^n$.

Definition 1.1. A PCA (IPS) is attractive if whenever $\eta \preccurlyeq \delta$, then $c(x, \eta) \le c(x, \delta)$ if $\eta(x) = \delta(x) = 0$ and $c(x, \eta) \ge c(x, \delta)$ if $\eta(x) = \delta(x) = 1$.

Heuristically, one has that 0's attract 0's and 1's attract 1's. There is another characterization of attractiveness which will be useful for us later. We say that a function f from X to $\mathbf{R}$ is increasing if $\eta \preccurlyeq \delta$ implies that $f(\eta) \le f(\delta)$. Let $\mathscr{M}$ denote the collection of increasing continuous functions on X. This then allows us to place a natural partial order on $P(X)$.

Definition 1.2. $\nu \preccurlyeq \mu$ if $\int_X f \, d\nu \le \int_X f \, d\mu$ for all functions f in $\mathscr{M}$.

The fact that $\preccurlyeq$ is a partial order on $P(X)$ is easily verified, as is the continuity of $\preccurlyeq$ with respect to the weak topology in that $\nu_n \preccurlyeq \mu_n$ for all n, $\nu_n \to \nu$ and $\mu_n \to \mu$ imply that $\nu \preccurlyeq \mu$.

We now present an alternative definition of attractiveness in the following proposition [see Liggett (1985) for the continuous time case].

Proposition 1.3. *A PCA (IPS) is attractive if and only if $\nu \preccurlyeq \mu$ implies $T\nu \preccurlyeq T\mu (T^t\nu \preccurlyeq T^t\mu$ for all $t)$.*

In the attractive context, there are two distinguished stationary distributions which might reduce to the same one. The next proposition deals with one of these.

PROPOSITION 1.4. $\lim_{n\to\infty} T^n\delta_0$ *exists and is a stationary distribution where* δ_0 *denotes the unit point mass at the configuration of all* 0*'s. Furthermore, this limiting distribution is smaller than any other stationary distribution with respect to the partial order defined in Definition* 1.2.

The analogous result holds when 0 is replaced by 1 or if discrete time is replaced by continuous time. This theorem is proved in Liggett (1985) in continuous time. The discrete time proof is analogous. We denote by $\underline{\nu}$ the stationary distribution whose existence is guaranteed by the above proposition. Similarly, if we start with the configuration of all 1's, we denote the limiting stationary distribution by $\bar{\nu}$. This notation is used in both discrete and continuous time. We point out that the main reason that one can obtain strong results in the attractive case is that the set $K = \{(\eta, \delta): \eta \preccurlyeq \delta\}$ is invariant for the *basic* coupling, which is defined in Section 3. The reason why we take $\frac{1}{2}$ as a uniform bound on the spin rates is that Propositions 1.3 and 1.4 become false otherwise, which one can see by taking $c(x, \eta) \equiv 1$.

In Section 2, we give all necessary definitions and background concerning the $\bar{d}$-metric, the notion of isomorphic processes and Bernoulli shifts. In Section 3, we define the *basic* couplings which we use throughout the paper. In Section 4, using $\underline{\nu}$ as our stationary distribution, we show that the corresponding stationary process viewed as an X-valued process indexed by $\mathbf{Z}$ is isomorphic (in the sense of ergodic theory) to an independent process also indexed by $\mathbf{Z}$. If the system is also assumed to be translation invariant in the obvious sense, we further show that this stationary process, viewed as a $\{0,1\}$-valued process indexed by $\mathbf{Z}^n \times \mathbf{Z}$ (space-time), is isomorphic to an independent process also indexed by $\mathbf{Z}^n \times \mathbf{Z}$, a much stronger ergodic property. Processes isomorphic to independent processes are called Bernoulli shifts. Finally, in Section 5, we extend all of these results to continuous time.

2. The $\bar{d}$-metric, isomorphism and Bernoulli shifts. We first introduce the $\bar{d}$-metric and some associated theorems which we will use. In certain contexts, this is also called the Vaserstein distance. Throughout this paper, $P(Y)$ will denote the set of probability measures on the space Y.

DEFINITION 2.1. If $\mu, \nu \in P(\{0, \dots, k-1\}^N)$ with N finite, then

$$\bar{d}^N(\mu, \nu) = \inf\left\{E^m\left(\frac{1}{N}\sum_{i=1}^{N} I_{\{x_i \neq y_i\}}\right)\right\},$$

the infimum being taken over all couplings m of μ and ν, where E^m denotes expectation with respect to m, I_A denotes the indicator function of A and a typical element of $\{0, \dots, k-1\}^N \times \{0, \dots, k-1\}^N$ is denoted by $(\{x_i\}_{i=1}^N, \{y_i\}_{i=1}^N)$.

A coupling m of μ and μ is a measure on $\{0, \dots, k-1\}^N \times \{0, \dots, k-1\}^N$ whose first and second marginals (projections) are μ and ν, respectively. The right side without the inf, which we denote by $\bar{d}_m^N(\mu, \nu)$, simply measures the expected percentage of errors with respect to the coupling m. We sometimes omit N in the notation when no confusion will arise.

DEFINITION 2.2. If $\mu, \nu \in P(\{0, \dots, k-1\}^{\mathbf{Z}^n})$ are $\mathbf{Z}^n$-invariant,

$$\bar{d}(\mu, \nu) = \lim_{N \to \infty} \bar{d}^{N^n}(\mu_N, \nu_N),$$

where σ_N is the projection of σ onto $\{0, 1, \dots, k-1\}^{\{1, 2, \dots, N\}^n}$ obtained by only considering the elements $\{1, 2, \dots, N\}^n \subset \mathbf{Z}^n$.

Here we are just projecting the measure down to a box in the lattice $\mathbf{Z}^n$ with side lengths N. For $n = 1$, it is shown in Ornstein (1974) that this limit exists, that $\bar{d}$ is a metric and that this metric can also be expressed as $\inf\{P^m\{(\eta, \delta): \eta(0) \neq \delta(0)\}\}$, the infimum being taken over all translation invariant couplings m of μ and ν. The general case is proved similarly.

Before discussing the notion of a Bernoulli shift, we introduce a few basic notions from ergodic theory.

DEFINITION 2.3. An *abstract dynamical system* is a quadruple $(\Omega, \mathscr{A}, \mu, \pi_G)$ where Ω is a set, $\mathscr{A}$ is a σ-field of subsets of Ω, μ is a probability measure on $\mathscr{A}$ and π_G is a group action of G on Ω by bijective bimeasurable measure-preserving transformations.

In this paper, G will be $\mathbf{Z}^n$ or $\mathbf{R}^n$ for some n. All probability spaces we consider will be Lebesgue spaces. A stationary process with index set $\mathbf{Z}^n$ and metric state space X is an example of a dynamical system, where Ω is $X^{\mathbf{Z}^n}$, $\mathscr{A}$ is the collection of Borel sets of Ω, μ is the measure on Ω corresponding to the joint distribution of the process and $\mathbf{Z}^n$ acts canonically on $X^{\mathbf{Z}^n}$. (The stationarity of the process is equivalent to this action being measure-preserving). In this paper, stationary processes are always viewed as dynamical systems in this way, and except for Section 5 these are the only dynamical systems that we consider. It is important to keep in mind that the group acting on a stationary process viewed as a dynamical system is the index set for the stationary process.

DEFINITION 2.4. $(\Omega, \mathscr{A}, \mu, \pi_G)$ is *isomorphic* to $(\Omega', \mathscr{B}, \nu, \Psi_G)$ if there are G-invariant measurable sets A contained in Ω and B contained in Ω' each of measure 1 such that for all g, π_g and Ψ_g are bijective when restricted to these sets and such that there exists a bimeasurable measure-preserving mapping f from A to B so that $f(\pi_g(x)) = \Psi_g(f(x))$ for all g in G and x in A.

In order to distinguish between dynamical systems, there are a number of ergodic theoretic properties, all of which are isomorphism invariants. We give here the definitions of ergodicity, mixing and Bernoulli.

Definition 2.5. A dynamical system $(\Omega, \mathscr{A}, \mu, \pi_G)$ is *ergodic* if whenever $\pi_g A = A$ for all g in G where A is measurable, then $\mu(A) = 0$ or 1.

Definition 2.6. A dynamical system $(\Omega, \mathscr{A}, \mu, \pi_G)$ is *mixing* if for all A and B in $\mathscr{A}$, $\lim_{g \to \infty} \mu(\pi_g(A) \cap B) = \mu(A)\mu(B)$.

Definition 2.7. $(\Omega, \mathscr{A}, \mu, \pi_{\mathbf{Z}^n})$ is *Bernoulli* if it is isomorphic to $(W^{\mathbf{Z}^n}, \mathscr{B}, p, \pi_{\mathbf{Z}^n})$ for some Lebesgue space W, where $\mathscr{B}$ is the canonical σ-field on the product space, p is product measure and $\pi_{\mathbf{Z}^n}$ is the canonical action of $\mathbf{Z}^n$ on $W^{\mathbf{Z}^n}$.

We sometimes call a system which is Bernoulli a Bernoulli shift and note that this simply means that it is isomorphic to an independent process. Note that we have only defined Bernoulli for $\mathbf{Z}^n$-actions. In Walters (1975), it is shown that for $\mathbf{Z}$-actions, Bernoulli implies mixing which in turn implies ergodicity. These implications are still true for $\mathbf{Z}^n$-actions.

We state an important theorem concerning Bernoulli shifts. In order to do this, we introduce a few preliminary notions.

First, we place a linear order $\prec^n$ on $\mathbf{Z}^n$. This ordering is defined by induction on n and can be thought of as a backwards dictionary ordering. If $n = 1$, we use the natural linear order $<$. For $n \geq 2$, we define $\prec^n$ by induction via

$$(a_1, \dots, a_n) \prec^n (b_1, \dots, b_n)$$

if

$$a_n < b_n \quad \text{or} \quad \left(a_n = b_n \text{ and } (a_1, \dots, a_{n-1}) \prec^{n-1} (b_1, \dots, b_{n-1})\right).$$

This clearly defines a linear ordering. If $a, b \in \mathbf{Z}$ with $a \leq b$, we let $[a, b]$ denote $\{y \in \mathbf{Z}\colon a \leq y \leq b\}$ and call this an *interval*. If $x \in \mathbf{Z}^n$, we let

$$\text{Past}(x) = \{y \in \mathbf{Z}^n\colon y \prec^n x\}.$$

If $R \subset \mathbf{Z}^n$ is a product of intervals, which we call a *rectangle*, we define the Past of R. We do this as follows. We tile the lattice $\mathbf{Z}^n$ by R and disjoint translates of R. There is then a natural mapping f from $\mathbf{Z}^n$ onto itself which maps each translate of R to a single point and sends R to the origin. We are just collapsing each translate of R to a point so that the relative position of two translates of R corresponds to the relative position of their images under f. The Past of R, denoted Past(R), is then defined to be $f^{-1}(\text{Past}(0))$, where Past(0) is defined above.

If $\{l_1, \dots, l_m\} \subset \mathbf{Z}^n$, we let $\sigma(l_1, \dots, l_m)$ denote the sub σ-field on $\{0, \dots, k-1\}^{\mathbf{Z}^n}$ generated by these m points and let $\mathscr{A}(\sigma(l_1, \dots, l_m))$ denote the collection of k^m atoms generating this sub σ-field.

DEFINITION 2.8. A $\mathbf{Z}^n$-invariant measure $\mu \in P(\{0,\dots,k-1\}^{\mathbf{Z}^n})$ is *very weak Bernoulli* (VWB) if $\forall\ \varepsilon > 0$ there is a rectangle $R \subset \mathbf{Z}^n$ such that if $\{l_1,\dots,l_m\} \subset \mathrm{Past}(R)$, then

$$\bar{d}(\mu|_R, \mu|_R/A) < \varepsilon$$

for all $A \in \mathscr{A}(\sigma(l_1,\dots,l_m))$ except for ε portion as measured by μ.

Here $\mu|_R$ denotes the measure on $\{0,\dots,k-1\}^R$ obtained by projecting μ onto $R \subset \mathbf{Z}^n$, $\mu|_R/A$ means $\mu|_R$ conditioned on A and the previous definition means that the union of the atoms A where the previous inequality fails has μ-measure $\leq \varepsilon$. We remark that the intervals whose product is R are not required to have the same length.

Before discussing the next object of study, we need to discuss the notion of entropy, which is an isomorphism invariant. Although we give the necessary definitions here, we refer the reader to Walters (1975) for a complete discussion.

DEFINITION 2.9. If $\mathscr{P} = \{P_1,\dots,P_s\}$ and $\mathscr{Q} = \{Q_1,\dots,Q_t\}$ are two finite partitions of a measure space $(X, \mathscr{B}, m)$, then the *conditional entropy of $\mathscr{P}$ given $\mathscr{Q}$*, denoted $H(\mathscr{P}\,|\,\mathscr{Q})$, is

$$-\sum_{i,j} m(P_i \cap Q_j)\log\frac{m(P_i \cap Q_j)}{m(Q_j)},$$

where log is the natural logarithm.

DEFINITION 2.10. If μ is a $\mathbf{Z}^n$-invariant measure on $P(\{0,\dots,k-1\}^{\mathbf{Z}^n})$ or equivalently a stationary process with state space $\{0,\dots,k-1\}$ and index set $\mathbf{Z}^n$, then the *entropy* of μ denoted by $H(\mu)$ is $\lim_{m\to\infty} H(\mathscr{P}|\mathscr{Q}_m)$, where $\mathscr{P}$ is the canonical partition of $\{0,\dots,k-1\}^{\mathbf{Z}^n}$ into k sets corresponding to the 0th coordinate, $\mathscr{Q}_m$ is the partition of $\{0,\dots,k-1\}^{\mathbf{Z}^n}$ into k^m sets corresponding to the coordinates $\{x_1,\dots,x_m\}$ in $\mathbf{Z}^n$ where Past(0) is enumerated $\{x_i\}_{i=1}^{\infty}$ and $H(\mathscr{P}\,|\,\mathscr{Q}_m)$ denotes the conditional entropy of $\mathscr{P}$ given $\mathscr{Q}_m$.

It can be shown that Definition 2.10 is independent of the enumeration of Past(0).

DEFINITION 2.11. A $\mathbf{Z}^n$-invariant measure $\mu \in P(\{0,\dots,k-1\}^{\mathbf{Z}^n})$ is *finitely determined* (FD) if the following condition holds: $\mu_n \to \mu$ weakly and $H(\mu_n) \to H(\mu)$ together imply that $\mu_n \to_{\bar{d}} \mu$.

The following theorem plays a central role in this paper. For the case $n = 1$, this theorem is due to Ornstein (1974) and Ornstein and Weiss (1974). For general n, one should refer to Katznelson and Weiss (1972), Feldman (1980), Conze (1972) and Thouvenot (1972).

THEOREM 2.12. *If $\mu \in P(\{0,\dots,k-1\}^{\mathbf{Z}^n})$ is $\mathbf{Z}^n$-invariant, then the following are equivalent*:

(i) *μ is Bernoulli.*
(ii) *μ is VWB.*
(iii) *μ is FD.*

We now state some other facts from the theory of Bernoulli shifts which we shall need. The proofs of these can be found in Ornstein and Weiss (1987).

THEOREM 2.13. *Let $X = \{0,1\}^{\mathbf{Z}^n}$ and μ be a measure on $X^{\mathbf{Z}}$ which is $\mathbf{Z}$-invariant. If $((\{0,1\}^{\{x_1,\dots,x_l\}})^{\mathbf{Z}}, \mu^{x_1,\dots,x_l}, \mathbf{Z})$ is Bernoulli for all l, where $\mu^{x_1,\dots,x_l}$ is the projection of μ onto $(\{0,1\}^{\{x_1,\dots,x_l\}})^{\mathbf{Z}}$ and $\mathbf{Z}^n$ is enumerated as $\{x_i\}_{i=1}^{\infty}$, then $(X^{\mathbf{Z}}, \mu, \mathbf{Z})$ is also Bernoulli.*

THEOREM 2.14. *The collection of Bernoulli measures and the collection of mixing measures are each closed in the $\bar{d}$-metric.*

THEOREM 2.15. *If $(X, \mathscr{B}, \mu, \phi_G)$ is a Bernoulli system and H is a subgroup of G, then $(X, \mathscr{B}, \mu, \phi_H)$ is a Bernoulli system, where ϕ_H is the restriction of the group action ϕ_G to H.*

Last, we state a general theorem about Markov processes which we shall need. This can be found in Rosenblatt (1971). First, a discrete time Markov kernel on a compact metric state space (X, d) with Borel σ-field $\mathscr{B}$ is a function

$$p(\cdot,\cdot)\colon X \times \mathscr{B} \to [0,1]$$

such that $p(x,\cdot)$ is a probability measure on $\mathscr{B}$ for each $x \in X$ and $p(\cdot, A)$ is measurable for each $A \in \mathscr{B}$. We assume that $p(x,\cdot)$ is continuous in x in the weak$*$ topology of measures. The natural operator on measures is then given by $T\mu(A) = \int_X p(x, A)\, d\mu(x)$. Again, μ is stationary if $T\mu = \mu$. We also have an operator on $C(X)$, the space of continuous functions on X, which we also denote by T given by $Tf(x) = \int_X f(y) p(x, dy)$.

THEOREM 2.16. *Let $\{p(x,\cdot)\}_{x\in X}$ be a discrete time Markov kernel with a compact metric state space X and let μ be an extreme point of the set of stationary distributions. Then*

$$\frac{1}{N}\sum_{i=0}^{N-1} T^i\delta_x \to \mu$$

weakly as $N \to \infty$ for μ a.e. x.

3. Basic couplings. In the theory of spin systems, coupling techniques have proved very fruitful. Many of these particular applications in continuous

TABLE 1
Basic coupling probabilities for k systems

Flip only	With probability
$\eta_{l_1}, \dots, \eta_{l_s}$	$c_{l_1}(x, \eta_{l_1})$
$\eta_{l_2}, \dots, \eta_{l_s}$	$c_{l_2}(x, \eta_{l_2}) - c_{l_1}(x, \eta_{l_1})$
$\eta_{l_3}, \dots, \eta_{l_s}$	$c_{l_3}(x, \eta_{l_3}) - c_{l_2}(x, \eta_{l_2})$
$\vdots$	$\vdots$
η_{l_s}	$c_{l_s}(x, \eta_{l_s}) - c_{l_{s-1}}(x, \eta_{l_{s-1}})$
$\eta_{l_{s+1}}, \dots, \eta_{l_k}$	$c_{l_{s+1}}(x, \eta_{l_{s+1}})$
$\eta_{l_{s+2}}, \dots, \eta_{l_k}$	$c_{l_{s+2}}(x, \eta_{l_{s+2}}) - c_{l_{s+1}}(x, \eta_{l_{s+1}})$
$\vdots$	$\vdots$
η_{l_k}	$c_{l_k}(x, \eta_{l_k}) - c_{l_{k-1}}(x, \eta_{l_{k-1}})$

time can be found in Liggett (1985). In particular, the *Vaserstein* or *basic* coupling of any two continuous time IPS is introduced there.

We introduce the discrete time analogue for k systems which we also call the basic coupling. We assume that we have k sets of spin rates $\{c_i(x, \eta)\}_{i=1}^k$. We wish to couple these k systems in a natural way and obtain a Markov kernel with state space X^k whose k marginals are the Markov kernels for the respective individual systems. The coupling is constructed as follows. What we do will become clear if one keeps in mind that we are trying to push the processes together as much as possible.

We assume that we are in the state $(\eta_1, \dots, \eta_k)$. We describe how to proceed at lattice point x and once this is done, all lattice points act independently. We break up the η_i's, $1 \le i \le k$, into two sets depending upon their value at x. Within each of the two sets, we reorder the η_i's in increasing order with respect to the values $\{c_i(x, \eta_i)\}_{i=1}^k$. This gives us two sequences $(\eta_{l_1}, \dots, \eta_{l_s})$ and $(\eta_{l_{s+1}}, \dots, \eta_{l_k})$, where the first half consists of those configurations which are in state 0 at x, the second half consists of those configurations which are in state 1 at x, and where both sequences are ordered according to their respective spin rate at x. We then proceed at lattice point x according to Table 1. The sum of the numbers in the right column of this table is at most 1. We flip no configuration at x with whatever remaining probability there is. It is clear that this is a coupling of the individual k systems. Furthermore, one can project this Markov kernel down to any subset of the k systems and obtain the coupling, as just defined, for this subset of the k systems. We call this our basic coupling and denote the corresponding Markov operator on $P(X \times X)$ or on $C(X \times X)$ by $\tilde{T}$.

4. Space-time Bernoullicity in discrete time. Before proving space-time Bernoullicity in the discrete time case when either $\underline{\nu}$ or $\bar{\nu}$ is used, we prove Bernoullicity under the time evolution. There is no translation invariance assumption here. The resulting measures on $X^{\mathbf{Z}}$ for the corresponding stationary processes are denoted by $\underline{\tilde{\nu}}$ and $\tilde{\bar{\nu}}$, respectively. Due to symmetry, we only consider $\underline{\tilde{\nu}}$ throughout this paper.

THEOREM 4.1. *If the spin rates are attractive, then* $(X^{\mathbf{Z}}, \mathscr{B}, \underline{\tilde{\nu}}, \mathbf{Z})$ *is Bernoulli, where* $\mathscr{B}$ *is the natural Borel structure on* $X^{\mathbf{Z}}$ *and where* $\mathbf{Z}$ *acts canonically on* $X^{\mathbf{Z}}$.

PROOF. Fix finitely many lattice points $\{x_1, \dots, x_l\} \subset \mathbf{Z}^n$. By Theorem 2.13, in order to prove Theorem 4.1, it suffices to prove the Bernoullicity of the process obtained by projecting onto lattice points $\{x_1, \dots, x_l\} \subset \mathbf{Z}^n$. This is just a measure $\underline{\tilde{\nu}}^{x_1, \dots, x_l}$ on $(\{0,1\}^l)^{\mathbf{Z}}$ which is $\mathbf{Z}$-invariant.

Using Theorem 2.12, it suffices to show that $\underline{\tilde{\nu}}^{x_1, \dots, x_l}$ as a process taking on values in $\{0,1\}^l$ with index set $\mathbf{Z}$ has the VWB property. Let $Y_i^{\{x_1, \dots, x_l\}}$ be the l-tuple in $\{0,1\}^l$ corresponding to the values of the configuration at lattice points $x_1, \dots, x_l$ at time i. $\underline{\tilde{\nu}}^{x_1, \dots, x_l}$ is therefore the joint distribution of the $Y_i^{\{x_1, \dots, x_l\}}$'s. To demonstrate the VWB property, it is useful to consider the entire spin system since $\underline{\tilde{\nu}}^{x_1, \dots, x_l}$ is non-Markovian and therefore difficult to handle.

We isolate the heart of the argument into the following lemma which we prove afterward. Here $\{Y_i^{\{x_1, \dots, x_l\}, \eta}\}_{i \geq 0}$ denotes the process $\{Y_i^{\{x_1, \dots, x_l\}}\}_{i \geq 0}$ conditioned on η at time 0.

LEMMA 4.2. *If the spin rates are attractive, then given* $\varepsilon > 0$, *there is an* N *and a set* $A \subseteq X$ *of* $\underline{\nu}$*-measure at least* $1 - \varepsilon$ *so that for all* $n \geq N$ *and for all* $\eta \in A$,

$$\bar{d}\left(\left\{Y_i^{\{x_1, \dots, x_l\}}\right\}_{i=1}^n, \left\{Y_i^{\{x_1, \dots, x_l\}, \eta}\right\}_{i=1}^n\right) \leq \varepsilon.$$

(Note that with regard to the $\bar{d}$-metric, we are viewing these processes as $\{0,1\}^l$-valued with index set $[1, n]$.)

We show how this lemma implies the VWB property. Let $\varepsilon > 0$. Choose N and A as in Lemma 4.2 using ε^2 instead of ε. Therefore $\underline{\nu}(A) \geq 1 - \varepsilon^2$ and given $\eta \in A$, there is a measure P^η on $(\{0,1\}^l)^N \times (\{0,1\}^l)^N$ which is a coupling of

$$\left\{Y_i^{\{x_1, \dots, x_l\}}\right\}_{i=1}^N \quad \text{and} \quad \left\{Y_i^{\{x_1, \dots, x_l\}, \eta}\right\}_{i=1}^N$$

satisfying

$$\bar{d}_{P^\eta}\left(\left\{Y_i^{\{x_1, \dots, x_l\}}\right\}_{i=1}^N, \left\{Y_i^{\{x_1, \dots, x_l\}, \eta}\right\}_{i=1}^N\right) \leq \varepsilon^2. \tag{4.1}$$

If $\eta \notin A$, let P^η be an arbitrary coupling of these two processes. Now if $\sigma_i \in \{0,1\}^l$ for $i = 0, \dots, m$, we obtain a measure $Q^{\{\sigma_0, \dots, \sigma_m\}}$ on X by considering the conditional distribution of Y_0 given

$$\left\{Y_{-i}^{\{x_1, \dots, x_l\}} = \sigma_i, i = 0, \dots, m\right\}.$$

Since

$$\sum_{(\sigma_0,\dots,\sigma_m)\in(\{0,1\}^l)^{m+1}} Q^{\{\sigma_0,\dots,\sigma_m\}}\tilde{\underline{\nu}}\{Y_{-i}^{\{x_1,\dots,x_l\}} = \sigma_i,\, i = 0,\dots,m\} = \underline{\nu}$$

and $\underline{\nu}(A) \geq 1 - \varepsilon^2$, it follows that

$$\tilde{\underline{\nu}}\left(\bigcup_{\{(\sigma_0,\dots,\sigma_m)\in\Theta\}} \{Y_{-i}^{\{x_1,\dots,x_l\}} = \sigma_i,\, i = 0,\dots,m\}\right) \leq \varepsilon,$$

where

$$\Theta = \left\{(\sigma_0,\dots,\sigma_m) \in \left(\{0,1\}^l\right)^{m+1} : Q^{\{\sigma_0,\dots,\sigma_m\}}(A) < 1 - \varepsilon\right\}.$$

We now complete the proof by showing that

$$Q^{\{\sigma_0,\dots,\sigma_m\}}(A) \geq 1 - \varepsilon$$

implies

$$\bar{d}\left(\{Y_i^{\{x_1,\dots,x_l\}}\}_{i=1}^N, \{Y_i^{\{x_1,\dots,x_l\}}\}_{i=1}^N / \{Y_{-i}^{\{x_1,\dots,x_l\}} = \sigma_i,\, i = 0,\dots,m\}\right) \leq 2\varepsilon.$$

This however is easy since

$$P^{\{\sigma_0,\dots,\sigma_m\}} = \int_X P^{\eta}\, dQ^{\{\sigma_0,\dots,\sigma_m\}}(\eta)$$

is a coupling of these two processes and furthermore

$$\bar{d}_{P^{\{\sigma_0,\dots,\sigma_m\}}}\left(\{Y_i^{\{x_1,\dots,x_l\}}\}_{i=1}^N, \{Y_i^{\{x_1,\dots,x_l\}}\}_{i=1}^N / \{Y_{-i}^{\{x_1,\dots,x_l\}} = \sigma_i,\, i = 0,\dots,m\}\right) \leq 2\varepsilon$$

by (4.1) and the fact that $Q^{\{\sigma_0,\dots,\sigma_m\}}$ gives A at least measure $1 - \varepsilon$. □

We now prove Lemma 4.2.

Proof of Lemma 4.2. To prove this lemma, we form the basic triple coupling of our process with initial distribution $\delta_0 \times \delta_\eta \times \underline{\nu}$, where η is now arbitrary. Let P^η denote the resulting measure on $(X \times X \times X)^{\mathbf{N}}$, where $\mathbf{N}$ are the nonnegative integers. Let Y_n^0, Y_n^η and Y_n denote the three marginal processes. If superscripts $\{x_1,\dots,x_l\}$ are attached to these processes, then this modification refers to the process projected onto lattice points $\{x_1,\dots,x_l\}$.

Note that the set

$$K = \{(\delta,\gamma,\sigma): \delta \preccurlyeq \gamma, \delta \preccurlyeq \sigma\}$$

is invariant for the triple process. Fix $x \in \mathbf{Z}^n$. The invariance of K together with the fact that $T^n\delta_0 \to \underline{\nu}$ weakly as $n \to \infty$ implies that for all η in X,

$$P^\eta\{Y_n^0(x) \neq Y_n(x)\} \to 0 \quad \text{as } n \to \infty.$$

From this and the invariance of K again, it follows that

$$(4.2)\qquad P^\eta\{Y_n^\eta(x) = 0, Y_n(x) = 1\} \to 0 \quad \text{as } n \to \infty \qquad \forall\, \eta \in X,$$

since $\{Y_n^\eta(x) = 0, Y_n(x) = 1\} \subset \{Y_n^0(x) \neq Y_n(x)\}$. (4.2) then implies that

$$\limsup_{n\to\infty} P^\eta\{Y_n^\eta(x) = 0\} \le \underline{\nu}\{\eta\colon \eta(x) = 0\}.$$

On the other hand, $\underline{\nu}$ is an extreme point of the set of stationary distributions. See Liggett (1985) for the continuous time proof of this fact while the discrete time proof is analogous. Therefore it follows from Theorem 2.24 that

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n\delta_\eta \to \underline{\nu} \quad \text{weakly as } N \to \infty$$

for $\underline{\nu}$ a.e. η. Let $\tilde{X}$ denote this set of full $\underline{\nu}$-measure. Hence for all $\eta \in \tilde{X}$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} P^\eta\{Y_n^\eta(x) = 0\} = \underline{\nu}\{\eta\colon \eta(x) = 0\}.$$

Next, in general, if $\{a_n\}_{n\ge 1}$ and a are all nonnegative with $\limsup_{n\to\infty} a_n \le a$ and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} a_n = a,$$

it follows that $a_n \to a$ on a set of density 1. Therefore, for all $\eta \in \tilde{X}$,

$$P^\eta\{Y_n^\eta(x) = 0\} \to \underline{\nu}\{\eta\colon \eta(x) = 0\} \quad \text{as } n \to \infty \text{ on a set of density } 1.$$

(The set of density 1 may depend upon η.) This last fact together with (4.2) implies that for all $\eta \in \tilde{X}$,

$$P^\eta\{Y_n^\eta(x) \neq Y_n(x)\} \to 0 \quad \text{as } n \to \infty \text{ on a set of density } 1,$$

from which one deduces that for all $\eta \in \tilde{X}$,

$$\frac{1}{N}\sum_{n=1}^{N} P^\eta\{Y_n^\eta(x) \neq Y_n(x)\} \to 0 \quad \text{as } N \to \infty.$$

Since $x \in \mathbf{Z}^n$ was arbitrary and l is fixed, this then implies that

$$\text{(4.3)}\quad \frac{1}{N}\sum_{n=1}^{N} P^\eta\{Y_n^{\{x_1,\dots,x_l\},\eta} \neq Y_n^{\{x_1,\dots,x_l\}}\} \to 0 \quad \text{as } N \to \infty \text{ for all } \eta \in \tilde{X}.$$

Since $\underline{\nu}(\tilde{X}) = 1$, the lemma follows with the projection of P^η being the desired coupling. □

Under the assumption of translation invariance, we now prove the space-time Bernoullicity of $\tilde{\underline{\nu}}$ in discrete time. To prove the Bernoullicity, we first deal with finite range spin rates and then extend to infinite range spin rates. [By finite range spin rates, we of course mean that $c(0, \eta)$ depends on η only through some fixed finite number of coordinates of η.] We prove this by showing that $\tilde{\underline{\nu}}$ has the VWB property. We introduce a number of lemmas throughout the argument in order to make it more palatable. As always,

$\{Y_i(x)\}_{(x,i)\in \mathbf{Z}^n\times\mathbf{Z}}$ denotes the stationary process with stationary distribution $\underline{\nu}$ and $\{Y_i^\eta(x)\}_{x\in\mathbf{Z}^n, i\geq 0}$ denotes the process with initial configuration η.

We verify the VWB property by showing the slightly stronger following condition:

For every $\varepsilon > 0$, there is a box $B_{a,b} \subset \mathbf{Z}^{n+1}$ of the form $\prod_{i=1}^n[-a,a]\times[1,b]$ such that if

$$\{y_1,\dots,y_r\}\subset (B_{a,b})^c\cap\{\omega\in\mathbf{Z}^{n+1}\colon \omega_{n+1}\leq b\},$$

the following holds: For all but $\sqrt{\varepsilon}$ portion of the atoms A in $\mathscr{A}(\sigma(y_1,\dots,y_r))$ (with respect to the probability measure $\tilde{\underline{\nu}}$),

$$\bar{d}\big(\{Y_i(x)\}_{(x,i)\in B_{a,b}},\{Y_i(x)\}_{(x,i)\in B_{a,b}}/A\big) < 10\varepsilon^{1/4}.$$

(Note that with regard to the $\bar{d}$-metric, we are now viewing these processes as $\{0,1\}$-valued with index set $B_{a,b}$ as opposed to $\{0,1\}^l$-valued with index set $[1,n]$ as we did before.)

The outline of the proof is as follows. One first proves this when $\{y_1,\dots,y_r\}$ all have their time coordinate being 0. This is the heart of the argument. Here one actually conditions on the entire configuration at time 0. Lemma 4.2 gives us that for most η the measure P^η on $(X\times X)^{\mathbf{N}}$ corresponding to the basic coupling couples Y_i^η and Y_i well at lattice point 0 for most i if one runs the process sufficiently long. Using Fubini's theorem, this then implies that most times i are good in that P^η couples Y_i^η and Y_i well at lattice point 0 for most η. Once one has this, one averages all the couplings P^η over η with respect to $\underline{\nu}$ to obtain a measure P on $(X\times X)^{\mathbf{N}}$. For each $i\in\mathbf{N}$, P yields a coupling of $\underline{\nu}$ with itself. One then shows that $\underline{\nu}$ is mixing and that $\tilde{T}$ takes ergodic measures to ergodic measures. (The only places where attractiveness is used are in the invoking of Lemma 4.2 and in showing that $\underline{\nu}$ is mixing.) One then concludes that these couplings of $\underline{\nu}$ with itself induced by P are ergodic. One can then apply the ergodic theorem (to the spatial action) to conclude that with respect to P, the two processes are at good times coupled well at most lattice points in arbitrarily large spatial boxes. One can then leave the averaged process P and conclude that at good times, for $\underline{\nu}$ most η, the basic coupling couples Y_i^η and Y_i well at most lattice points in arbitrarily large spatial boxes. This will take care of the case where $\{y_1,\dots,y_r\}$ all have their time coordinate being 0. Once one knows how long (timewise) one must wait to get a good coupling, one then chooses a exceedingly large relative to b and uses the finite range assumption to show that further conditioning on lattice points in $(B_{a,b})^c$ whose time coordinate is in $[1,b]$ has only a little effect. This is the basic outline of the proof. In verifying the VWB property, we first prove Proposition 4.3 which contains the heart of the argument.

PROPOSITION 4.3. *Given $\varepsilon>0$, there is an integer b and a set $E\subseteq X$ of $\underline{\nu}$-measure at most ε such that for all a sufficiently large,*

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i)\in B_{a,b}},\{Y_i(x)\}_{(x,i)\in B_{a,b}}\big)\leq 8\varepsilon^{1/4}$$

for all $\eta \notin E$, *where*

$$B_{a,b} = \prod_{i=1}^{n} [-a, a] \times [1, b].$$

PROOF. We already have a canonical way of coupling $\{Y_i^\eta\}_{i \geq 0}$ and $\{Y_i\}_{i \geq 0}$, namely, the basic coupling together with initial distribution $\delta_\eta \times \underline{\nu}$. Our goal is to show that this particular coupling is the desired $\bar{d}$-coupling. We let P^η denote the resulting measure on $(X \times X)^{\mathbf{N}}$. Given any $\varepsilon > 0$, Lemma 4.2 implies that there is an integer m and a set $A \subseteq X$ such that $\underline{\nu}(A) \geq 1 - \varepsilon$ and $\forall\ \eta \in A$,

$$E^{P_\eta}\left[\frac{1}{m} \sum_{i=1}^{m} I_{\{Y_i^\eta(0) \neq Y_i(0)\}}\right] \leq \varepsilon.$$

Letting u be uniform measure on $\{1, \dots, m\}$, one can apply Fubini's theorem to the space $(\{1, \dots, m\} \times X, u \times \underline{\nu})$ and the function $g(i, \eta) = P^\eta\{Y_i^\eta(0) \neq Y_i(0)\}$ to show that this implies that for all $k \in \{1, \dots, m\}$ except for at most $(2\varepsilon)^{1/2}$ portion, the following holds:

$$P^\eta\{Y_k^\eta(0) \neq Y_k(0)\} \leq (2\varepsilon)^{1/4}$$

for all η except for a set of $\underline{\nu}$-measure at most $(2\varepsilon)^{1/4}$. Let A_k be the set of η where the above inequality is satisfied and let S_m^ε denote the good set of $k \in \{1, \dots, m\}$ of percentage greater than or equal to $1 - (2\varepsilon)^{1/2}$.

Now let $P^{\eta, k}$ denote the measure on $X \times X$ obtained by projecting P^η onto the time k coordinate. As indicated in the outline of the proof, we want to average the couplings $\{P^\eta\}_{\eta \in X}$ with respect to $\underline{\nu}$. We therefore let

$$\mu_k = \int_X P^{\eta, k}\, d\underline{\nu}(\eta).$$

Clearly μ_k is a $\underline{\nu}$-$\underline{\nu}$ coupling which is translation invariant. Restricting to good times, we want to use the ergodic theorem to conclude from the fact that for most η we have a good coupling at lattice point 0 that we also have a good coupling at most lattice points for most η. We therefore naturally want that μ_k is spatially ergodic. This is the content of Lemma 4.4, whose proof we give later.

LEMMA 4.4. μ_k *is ergodic under the natural* $\mathbf{Z}^n$*-action for all* k.

Next, if $k \in S_m^\varepsilon$, then

$$\begin{aligned}\mu_k\{(\delta, \gamma) : \delta(0) \neq \gamma(0)\} &= \int_X P^\eta\{Y_k^\eta(0) \neq Y_k(0)\}\, d\underline{\nu}(\eta)\\ &\leq (2\varepsilon)^{1/4} + (2\varepsilon)^{1/4} \leq 4\varepsilon^{1/4},\end{aligned}$$

by integrating over the pieces A_k and $(A_k)^c$.

Applying the ergodic theorem to μ_k, we can choose a box $B_k \subset \mathbf{Z}^n$ (all of whose sides have equal length) sufficiently large so that the ergodic theorem kicks in to within ε on all but ε^2/m^2 of the space for the function $f(\delta,\gamma) = I_{\{\delta(0)\neq\gamma(0)\}}$ defined on $X\times X$. We can therefore find $\tilde{E}_k \subseteq X\times X$ such that $\mu_k(\tilde{E}_k) \le \varepsilon^2/m^2$ and if $(\zeta,\sigma)\notin \tilde{E}_k$, then

$$\left|\frac{1}{|B|}\sum_{x\in B} I_{\{\zeta(x)\neq\sigma(x)\}} - \mu_k\{(\delta,\gamma)\colon \delta(0)\neq\gamma(0)\}\right| \le \varepsilon$$

for all boxes $B\supseteq B_k$. If $k\in S_m^\varepsilon$, it follows that for $(\zeta,\sigma)\notin\tilde{E}_k$,

$$\frac{1}{|B|}\sum_{x\in B} I_{\{\zeta(x)\neq\sigma(x)\}} \le 4\varepsilon^{1/4}+\varepsilon \le 5\varepsilon^{1/4}$$

for all $B\supseteq B_k$.

Letting B_m^ε be the largest of the boxes B_k for $k\in S_m^\varepsilon$, we have that $\forall\, k\in S_m^\varepsilon$ $\forall\,(\zeta,\sigma)\notin\tilde{E}_k$ and $\forall\, B\supseteq B_m^\varepsilon$,

$$\frac{1}{|B|}\sum_{x\in B} I_{\{\zeta(x)\neq\sigma(x)\}} \le 5\varepsilon^{1/4}.$$

Let $k\in S_m^\varepsilon$. Since $\mu_k(\tilde{E}_k)\le\varepsilon^2/m^2$. It follows that

$$P^{\eta,k}\big(\tilde{E}_k\big) \le \frac{\varepsilon}{m}$$

for all $\eta\in X$ except for a set of at most $\underline{\nu}$-measure ε/m. Let E_k denote the set of η where this inequality fails and so

$$\underline{\nu}(E_k) \le \varepsilon/m.$$

Letting $E=\bigcup_{k\in S_m^\varepsilon} E_k$, we have that $\underline{\nu}(E)\le\varepsilon$.

Let a be sufficiently large so that $B_m^\varepsilon \subseteq \prod_{i=1}^n[-a,a]$ and let $b=m$. To prove the proposition, we show that $\forall\,\eta\notin E$,

$$E^{P^\eta}\left[\frac{1}{|B_{a,b}|}\sum_{(x,i)\in B_{a,b}} I_{\{Y_i^\eta(x)\neq Y_i(x)\}}\right] \le 8\varepsilon^{1/4}.$$

By the previous equation, for $\eta\notin E$,

$$P^{\eta,k}\big(\tilde{E}_k\big) \le \frac{\varepsilon}{m}\forall\, k\in S_m^\varepsilon.$$

We denote $\prod_{i=1}^n[-a,a]$ by B_a for convenience. It follows that for $k\in S_m^\varepsilon$,

$$\begin{aligned}
&E^{P^{\eta,k}}\left[\frac{1}{|B_a|}\sum_{x\in B_a} I_{\{Y_k^\eta(x)\neq Y_k(x)\}}\right]\\
&\quad= E^{P^{\eta,k}}\left[I_{(\tilde{E}_k)^c}\frac{1}{|B_a|}\sum_{x\in B_a} I_{\{Y_k^\eta(x)\neq Y_k(x)\}}\right] + E^{P^{\eta,k}}\left[I_{\tilde{E}_k}\frac{1}{|B_a|}\sum_{x\in B_a} I_{\{Y_k^\eta(x)\neq Y_k(x)\}}\right]\\
&\quad\le 5\varepsilon^{1/4}+\frac{\varepsilon}{m} \le 6\varepsilon^{1/4}.
\end{aligned}$$

It now follows that

$$E^{P^{\eta}}\left[\frac{\sum_{(x,i)\in B_{a,b}} I_{\{Y_i^{\eta}(x)\neq Y_i(x)\}}}{|B_{a,b}|}\right]$$

$$= \frac{\sum_{i=1}^{m} E^{P^{\eta,i}}\left[\sum_{x\in B_a} I_{\{Y_i^{\eta}(x)\neq Y_i(x)\}}\right]}{|B_{a,b}|}$$

$$= \frac{\sum_{\substack{i=1\\ i\in Sm^{\varepsilon}}}^{m} E^{P^{\eta,i}}\left[\sum_{x\in B_a} I_{\{Y_i^{\eta}(x)\neq Y_i(x)\}}\right]}{|B_{a,b}|} + \frac{\sum_{\substack{i=1\\ i\notin Sm^{\varepsilon}}}^{m} E^{P^{\eta,i}}\left[\sum_{x\in B_a} I_{\{Y_i^{\eta}(x)\neq Y_i(x)\}}\right]}{|B_{a,b}|}$$

$$\leq \frac{\sum_{\substack{i=1\\ i\in Sm^{\varepsilon}}}^{m} |B_a| 6\varepsilon^{1/4}}{|B_{a,b}|} + \frac{\sum_{\substack{i=1\\ i\notin Sm^{\varepsilon}}}^{m} |B_a|}{|B_{a,b}|}$$

$$\leq \frac{6\varepsilon^{1/4} m |B_a|}{|B_{a,b}|} + \frac{(2\varepsilon)^{1/2} m |B_a|}{|B_{a,b}|} \quad \left[\text{since } |S_m^{\varepsilon}| \geq \left(1 - (2\varepsilon)^{1/2}\right)m\right]$$

$$= 6\varepsilon^{1/4} + (2\varepsilon)^{1/2} \leq 8\varepsilon^{1/4}. \qquad \square$$

We now prove Lemma 4.4.

Proof of Lemma 4.4. We first need to show that $\underline{\nu}$ is mixing. This is proved in the continuous time case in the IPS literature. We give a simpler proof here based on the $\bar{d}$-metric. By coupling δ_0 with $\underline{\nu}$ in our basic coupling and letting time go to ∞, it is clear that $T^n\delta_0 \to_{\bar{d}} \underline{\nu}$ as $n \to \infty$. Since the set of mixing measures on $\{0,1\}^{\mathbf{Z}^n}$ is $\bar{d}$-closed by Theorem 2.14, it suffices to show that each $T^n\delta_0$ is mixing and hence to show that T takes mixing measures to mixing measures since δ_0 is mixing. This is proved in a standard way as follows [see Liggett (1985) for the continuous time version of an analogous result]. First note that $\mathbf{Z}^n$ acts canonically on X which induces an action on $C(X)$ which we denote by τ_y. For finite range spin rates, assuming ν is mixing and letting f and g depend on only finitely many lattice points,

$$\lim_{y\to\infty} \int_X (\tau_y(f))(\eta) g(\eta)\, dT\,\nu(\eta)$$

$$= \lim_{y\to\infty} \int_X T(\tau_y(f)g)(\eta)\, d\nu(\eta)$$

$$= \lim_{y\to\infty} \int_X T(\tau_y(f))(\eta) T(g)(\eta)\, d\nu(\eta)$$

[since f and g depend on only finitely many lattice points and T is finite range]

$$= \lim_{y\to\infty} \int_X \tau_y(T(f))(\eta) T(g)(\eta)\, d\nu(\eta) \quad \text{[by spatial invariance]}$$

$$= \int_X T(f)(\eta)\, d\nu(\eta) \int_X T(g)(\eta)\, d\nu(\eta) \quad \text{[as } \nu \text{ is mixing]}$$

$$= \int_X f(\eta)\, dT\nu(\eta) \int_X g(\eta)\, dT\,\nu(\eta).$$

Since the class of functions depending upon only finitely many lattice points is dense in $C(X)$, it follows that $T\nu$ is also mixing. We conclude that $\underline{\nu}$ is mixing.

To show that μ_k is ergodic, we show that μ_0 is ergodic and that $\tilde{T}$, the coupled process, preserves ergodic measures on $X \times X$. Once this is done, we need only check that $\tilde{T}\mu_k = \mu_{k+1}$ which is obtained as follows:

$$\tilde{T}\mu_k = \tilde{T} \int_X P^{\eta,k}\, d\underline{\nu}(\eta)$$

$$= \int_X \tilde{T}(P^{\eta,k})\, d\underline{\nu}(\eta) \quad \left[\text{as } \tilde{T} \text{ is a continuous operator}\right]$$

$$= \int_X P^{\eta,k+1}\, d\underline{\nu}(\eta) = \mu_{k+1}.$$

We next note that $\mu_0 = \underline{\nu} \times \underline{\nu}$ since

$$\mu_0(A \times B) = \int_X P^{\eta,0}(A \times B)\, d\underline{\nu}(\eta)$$

$$= \int_X (\delta_\eta \times \underline{\nu})(A \times B)\, d\underline{\nu}(\eta)$$

$$= \int_X \delta_\eta(A) \underline{\nu}(B)\, d\underline{\nu}(\eta)$$

$$= \left(\int_X \delta_\eta(A)\, d\underline{\nu}(\eta)\right) \underline{\nu}(B)$$

$$= \underline{\nu}(A)\underline{\nu}(B) = \underline{\nu} \times \underline{\nu}(A \times B).$$

Since $\underline{\nu}$ is mixing, it follows that $\mu_0 = \underline{\nu} \times \underline{\nu}$ is ergodic [see Walters (1975)].

Finally one shows that $\tilde{T}$ preserves ergodic measures in a way analogous to the proof that T preserves mixing measures and so we omit the proof which is in Liggett (1985) for continuous time. □

We are now in a position to prove our main theorem using what we have just proved together with the following lemma. This next lemma shows that the process conditioned on η at time 0 in $B_{a,b}$ is $\bar{d}$-close to itself if we further condition on elements in $\mathbf{Z}^{n+1}$ whose spatial coordinates do not lie in $B_a = \prod_{i=1}^n [-a, a]$ and whose time point is in $[1, b]$ providing a is sufficiently large relative to b. It is precisely in this lemma where the assumption of finite range comes into play.

LEMMA 4.5. *If the spin rates are finite range, then given any b and $\xi > 0$, one has that for all a sufficiently large, the following holds: If $\eta \in X$ and*

$$\{l_1, \dots, l_m\} \subseteq (B_a)^c \times [1, b],$$

then for all $A \in \mathscr{A}(\sigma(l_1, \dots, l_m))$,

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}, \{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}/A\big) \le \xi.$$

PROOF. Since the process has finite range, if the spin system starts in a given configuration, lattice points sufficiently far away from each other evolve completely independently for some finite amount of time. Hence for fixed b, there is a K such that if the spin system begins in any fixed configuration, then lattice points more than distance K away (in L^1 distance on $\mathbf{Z}^n$) evolve completely independently up until time b. Now choose a sufficiently large so that the points in B_a which are distance less than or equal to K from some point in $(B_a)^c$ have cardinality less than or equal to $\xi(2a + 1)^n$, that is, so that this set of points fills up at most ξ portion of the whole set B_a. Let $S \subseteq B_a$ be those points which are not in this set, that is, have distance greater than K from $(B_a)^c$. Then $|S|/(2a + 1)^n \ge 1 - \xi$.

Now given $\{l_1, \dots, l_m\} \subseteq (B_a)^c \times [1, b]$, the way we have chosen S implies that for all $A \in \mathscr{A}(\sigma(l_1, \dots, l_m))$, $\{Y_i^\eta(x)\}_{(x,i) \in S \times [1,b]}$ and $\{Y_i^\eta(x)\}_{(x,i) \in S \times [1,b]}/A$ have exactly the same joint distributions. Since

$$\frac{|S \times [1, b]|}{|B_{a,b}|} \ge 1 - \xi,$$

this clearly implies that

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}, \{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}/A\big) \le \xi. \qquad \square$$

THEOREM 4.6. *If the spin rates are attractive and of finite range, then*

$$\big(X^{\mathbf{Z}}, \mathscr{B}, \underline{\tilde{\nu}}, \mathbf{Z}^{n+1}\big)$$

is Bernoulli, where $\mathscr{B}$ is the natural Borel structure on $X^{\mathbf{Z}}$ and where $\mathbf{Z}^{n+1}$ acts canonically on $X^{\mathbf{Z}}$ which is identified with $\{0, 1\}^{\mathbf{Z}^{n+1}}$.

PROOF. We demonstrate the VWB property and then apply Theorem 2.12. Let $\varepsilon > 0$. By Proposition 4.3, there are integers a and b and a set $E \subseteq X$ with $\underline{\nu}(E) \le \varepsilon$ so that: If $a' \ge a$ and $\eta \notin E$, then

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i) \in B_{a',b}}, \{Y_i(x)\}_{(x,i) \in B_{a',b}}\big) \le 8\varepsilon^{1/4}.$$

By Lemma 4.5, we can choose a' sufficiently large (since b is fixed) so that if

$$\{l_1, \dots, l_m\} \subseteq (B_{a'})^c \times [1, b],$$

then for all $A \in \mathscr{A}(\sigma(l_1, \dots, l_m))$ and for all $\eta \in X$,

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i) \in B_{a',b}}, \{Y_i^\eta(x)\}_{(x,i) \in B_{a',b}}/A\big) \le \varepsilon.$$

Together this gives us

$$\bar{d}\left(\{Y_i^\eta(x)\}_{(x,i)\in B_{a',b}}/A, \{Y_i(x)\}_{(x,i)\in B_{a',b}}\right) \le 9\varepsilon^{1/4} \tag{4.4}$$

for $\eta \notin E$, $A \in \mathscr{A}(\sigma(l_1,\dots,l_m))$.

Now let

$$\{y_1,\dots,y_r\} \subseteq (B_{a,b})^c \cap \{\omega \in \mathbf{Z}^{n+1}: \omega_{n+1} \le b\}.$$

Partition $\{y_1,\dots,y_r\}$ into two sets $\{l_1,\dots,l_m\}$ and $\{l'_1,\dots,l_{m'}\}$, where $l_1,\dots,l_m$ all have their time coordinate positive and $l'_1,\dots,l_{m'}$ all have their time coordinate nonpositive. Then any element in $\mathscr{A}(\sigma(y_1,\dots,y_r))$ is $A_1 \cap A_2$, where $A_1 \in \mathscr{A}(\sigma(l_1,\dots,l_m))$ and $A_2 \in \mathscr{A}(\sigma(l'_1,\dots,l_{m'}))$.

Now we let Q^{A_2} denote the measure on X obtained by considering the conditional distribution of Y_0 given A_2. By (4.4), if $\eta \notin E$, there is a measure P^{η, A_1} on $\{0,1\}^{B_{a',b}} \times \{0,1\}^{B_{a',b}}$ which is a

$$\{Y_i^\eta(x)\}_{(x,i)\in B_{a',b}}/A_1 - \{Y_i(x)\}_{(x,i)\in B_{a',b}}$$

coupling which gives $\bar{d}$-distance $\le 9\varepsilon^{1/4}$. If $\eta \in E$, we let P^{η, A_1} be any

$$\{Y_i^\eta(x)\}_{(x,i)\in B_{a',b}}/A_1 - \{Y_i(x)\}_{(x,i)\in B_{a',b}}$$

coupling. It then follows that

$$P^{A_1, A_2} = \int_X P^{\eta, A_1}\, dQ^{A_2}(\eta)$$

is a coupling of

$$\{Y_i(x)\}_{(x,i)\in B_{a',b}}/(A_1 \cap A_2) \quad \text{and} \quad \{Y_i(x)\}_{(x,i)\in B_{a',b}}.$$

However, this is only a good coupling if Q^{A_2} does not give too much measure to E. In particular, if $Q^{A_2}(E) \le \sqrt{\varepsilon}$, then

$$\begin{aligned}\bar{d}_{P^{A_1,A_2}}&\left(\{Y_i(x)\}_{(x,i)\in B_{a',b}}/(A_1 \cap A_2), \{Y_i(x)\}_{(x,i)\in B_{a',b}}\right)\\ &\le 9\varepsilon^{1/4} + \sqrt{\varepsilon} < 10\varepsilon^{1/4}.\end{aligned}$$

Since

$$\sum_{A_2 \in \mathscr{A}(\sigma(l'_1,\dots,l'_{m'}))} Q^{A_2}\tilde{\underline{\nu}}(A_2) = \underline{\nu}$$

and $\underline{\nu}(E) \le \varepsilon$, it follows that

$$\tilde{\underline{\nu}}\left\{\bigcup_{\{A_2 \in \mathscr{A}(\sigma(l'_1,\dots,l'_{m'})):\ Q^{A_2}(E) \ge \sqrt{\varepsilon}\}} A_2\right\} \le \sqrt{\varepsilon}.$$

Hence for all $A \in \mathscr{A}(\sigma(y_1,\dots,y_r))$ except for $\sqrt{\varepsilon}$ portion with respect to $\tilde{\underline{\nu}}$,

$$\bar{d}\left(\{Y_i(x)\}_{(x,i)\in B_{a',b}}/A, \{Y_i(x)\}_{(x,i)\in B_{a',b}}\right) < 10\varepsilon^{1/4}.$$

This demonstrates the VWB property. □

We now extend this result to infinite range systems.

THEOREM 4.7. *If the spin rates are attractive, then* $(X^{\mathbf{Z}}, \mathscr{B}, \tilde{\underline{\nu}}, \mathbf{Z}^{n+1})$ *is Bernoulli, where* $\mathscr{B}$ *is the natural Borel structure on* $X^{\mathbf{Z}}$ *and where* $\mathbf{Z}^{n+1}$ *acts canonically on* $X^{\mathbf{Z}}$ *which is identified with* $\{0,1\}^{\mathbf{Z}^{n+1}}$.

PROOF. Let $\{c_L(x,\eta)\}$ denote the spin rates corresponding to modifying the original spin rates to make them finite range by cutting off their sight more than L units away. Formally, $c_L(0,\eta) = c(0,\eta_L)$, where

$$\eta_L(x) = \begin{cases} \eta(x) & \text{if } |x| \le L, \\ 0 & \text{if } |x| > L, \end{cases}$$

and then we take $c_L(x,\eta)$ to be $c_L(0,\tau_{-x}\eta)$. Here, $|x|$ denotes the L^1 norm on $\mathbf{Z}^n$ of x. Clearly the rates $\{c_L(x,\eta)\}$ are translation invariant, attractive and finite range. Let $\underline{\nu}^L$ denote the smallest stationary distribution for these rates and $\tilde{\underline{\nu}}^L$ the measure on $\{0,1\}^{\mathbf{Z}^{n+1}}$ corresponding to the resulting stationary process. Theorem 4.6 then tells us that $\tilde{\underline{\nu}}^L$ is Bernoulli. If we can show that $\tilde{\underline{\nu}}^L \to_{\bar{d}} \tilde{\underline{\nu}}$ as $L \to \infty$, we can then invoke Theorem 2.14 to conclude that $\tilde{\underline{\nu}}$ is also Bernoulli, as desired.

We note that it is clear that the $\{c_L(x,\eta)\}$ rates can be coupled below the $\{c(x,\eta)\}$ rates; that is,

$$K = \{(\eta,\delta)\colon \eta \preccurlyeq \delta\} \subseteq X \times X$$

is invariant for the basic coupling of the $\{c_L(x,\eta)\}$ and $\{c(x,\eta)\}$ rates. Now consider the basic coupling with initial distribution $\delta_0 \times \delta_0$. Since K is invariant, one can conclude that $\underline{\nu}^L \preccurlyeq \underline{\nu}$ by letting the coupled process run in time to ∞. Next, an easy argument shows that any weak limit of $\{\underline{\nu}^L\}$ must be invariant for the rates $\{c(x,\eta)\}$. Since $\underline{\nu}^L \preccurlyeq \underline{\nu}$ for all L it follows that $\underline{\nu}^L \to \underline{\nu}$ weakly as $L \to \infty$.

Now letting $\varepsilon > 0$, we have that for all large L,

$$|\underline{\nu}^L\{\eta\colon \eta(0) = 0\} - \underline{\nu}\{\eta\colon \eta(0) = 0\}| \le \varepsilon.$$

We claim that

$$\bar{d}(\tilde{\underline{\nu}}^L, \tilde{\underline{\nu}}) \le \varepsilon \quad \text{for all large } L.$$

Consider the coupled process $\tilde{T}_L$ for the rates $\{c_L(x,\eta)\}$ and $\{c(x,\eta)\}$ with initial distribution $\delta_0 \times \delta_0$. Taking a weak limit of

$$\frac{1}{N}\sum_{n=0}^{N-1} (\tilde{T}_L)^n(\delta_0 \times \delta_0)$$

yields a measure m on $X \times X$ which is stationary for the coupled process, is a $\underline{\nu}^L - \underline{\nu}$ coupling, has its support in K and is invariant under the natural $\mathbf{Z}^n$-action.

We next note that since $m(K) = 1$ and

$$|\underline{\nu}^L\{\eta\colon \eta(0) = 0\} - \underline{\nu}\{\eta\colon \eta(0) = (0)\}| \le \varepsilon,$$

it follows that

$$m\{(\eta,\delta):\eta(0)\neq\delta(0)\}\leq\varepsilon.$$

Since m is both translation invariant under the $\mathbf{Z}^n$-action and stationary for the coupled process, it follows that the stationary process

$$\tilde{m}\in P\Big((\{0,1\}\times\{0,1\})^{\mathbf{Z}^{n+1}}\Big)$$

induced from m is a coupling of $\tilde{\underline{\nu}}^L$ and $\tilde{\underline{\nu}}$ which is invariant under the entire $\mathbf{Z}^{n+1}$-action. Since $m\{(\eta,\delta)\colon \eta(0)\neq\delta(0)\}\leq\varepsilon$, it follows that

$$\bar{d}_{\tilde{m}}\big(\tilde{\underline{\nu}}^L,\tilde{\underline{\nu}}\big)\leq\varepsilon$$

as desired. Hence $\tilde{\underline{\nu}}^L\to_{\bar{d}}\tilde{\underline{\nu}}$ as $L\to\infty$. □

5. Space-time Bernoullicity in continuous time. Throughout this section, $D_X(\mathbf{R})$ is the space of right-continuous functions with left limits from $\mathbf{R}$ to X. If $\mathbf{R}$ is replaced by $[0,\infty)$, there is the so-called Skorohod metric on this space [see Ethier and Kurtz (1986) for details of this metric]. In this case, the Borel field generated by this metric can be shown to be [see Ethier and Kurtz (1986)] the same as the Borel field generated by the projection maps $f\to f(t)$ from $D_X([0,\infty))$ to X for $t\geq 0$. Therefore, by analogy, in dealing with $D_X(\mathbf{R})$, we put the Borel structure $\mathscr{B}$ on it that is generated by the above projections for $t\in\mathbf{R}$.

Our continuous time process clearly yields an action of $\mathbf{R}\times\mathbf{Z}^n$ on the space $D_X(\mathbf{R})$. Extending the definition of Bernoullicity to $\mathbf{R}^d\times\mathbf{Z}^n$-actions is done as follows. In general, an action of $\mathbf{R}^d\times\mathbf{Z}^n$ is called Bernoulli if the action of the restriction to $\mathbf{Z}^d\times\mathbf{Z}^n$ is Bernoulli as defined in Section 2. Moreover, for such actions, one still has that Bernoulli implies mixing which in turn implies ergodicity. We need a slight generalization of Theorem 2.13, Lemma 5.1, whose proof can also be found in Ornstein and Weiss (1987).

LEMMA 5.1. *Let $\mathscr{B}_m$ be the sub σ-field of $\mathscr{B}$ generated by time points $\{k/2^m\}_{k\in\mathbf{Z}}$. If $(D_X(\mathbf{R}),\mathscr{B}_m,\tilde{\underline{\nu}},\mathbf{Z}\times\mathbf{Z}^n)$ $((D_X(\mathbf{R}),\mathscr{B}_m,\tilde{\underline{\nu}},\mathbf{Z}))$ is Bernoulli for each m, then $(D_X(\mathbf{R}),\mathscr{B},\tilde{\underline{\nu}},\mathbf{Z}\times\mathbf{Z}^n)$ $((D_X(\mathbf{R}),\mathscr{B},\tilde{\underline{\nu}},\mathbf{Z}))$ is Bernoulli.*

We are now assuming the three conditions on our spin rates given in Section 1 to insure the existence of the continuous time process. Before proving space-time Bernoullicity in the continuous time case when the stationary distribution $\underline{\nu}$ is used, we discuss Bernoullicity under the time evolution where there is no translation invariance assumption. Using Lemma 5.1 and Theorem 2.15, it suffices to show that the continuous time process projected onto time points of the form $\{k/2^m\}_{k\in\mathbf{Z}}$ (which is just a measure on $X^{\mathbf{Z}}$) is Bernoulli for each m. Without loss of generality, we take $m=0$. One should note that this process which is Markovian is not a discrete time PCA and that the transition probabilities are much more complicated than for a PCA. Nonetheless, one can follow the proof of Theorem 4.1 almost verbatim to prove the following theorem, whose proof we therefore omit. One need only use the continuous time coupling [defined in Liggett (1985)] and make the observation that $\underline{\nu}$ is still an extreme point among the stationary distributions for the Markov process which is the projection of the continuous time process onto integer times.

THEOREM 5.2. *If the spin rates are attractive, then $(D_X(\mathbf{R}), \mathscr{B}, \tilde{\underline{\nu}}, \mathbf{R})$ is Bernoulli, where $D_X(\mathbf{R})$ and $\mathscr{B}$ are as defined above and where $\mathbf{R}$ acts canonically on $D_X(\mathbf{R})$.*

The proof of space-time Bernoullicity in continuous time is not just a trivial extension of the discrete time proof since Lemma 4.5 is no longer true as stated. We use ideas from first-passage percolation. We first prove space-time Bernoullicity for finite range spin rates and then extend to infinite range spin rates as in Theorem 4.7.

THEOREM 5.3. *If the spin rates are attractive and of finite range, then*

$$(D_X(\mathbf{R}), \mathscr{B}, \tilde{\underline{\nu}}, \mathbf{R} \times \mathbf{Z}^n)$$

is Bernoulli where $D_X(\mathbf{R})$ and $\mathscr{B}$ are as defined above and where $\mathbf{R} \times \mathbf{Z}^n$ acts canonically on $D_X(\mathbf{R})$.

PROOF. Our comments concerning the proof of Theorem 5.2 also hold here and so it suffices to show space-time Bernoullicity of the process projected onto integer time points.

We let $Y_i^\eta(x)$ and $Y_i(x)$ be defined as before. We let P^η denote the measure on $(X \times X)^{\mathbf{N}}$ obtained from projecting the continuous time basic coupling [defined in Liggett (1985)] with initial distribution $\delta_\eta \times \underline{\nu}$ onto integer times. We now state two propositions, the first of which we omit the proof of (it is proved exactly as is Proposition 4.3) and the second of which we prove later.

PROPOSITION 5.4. *Given $\varepsilon > 0$, there is an integer b and a set $E \subseteq X$ of $\underline{\nu}$-measure at most ε such that for all a sufficiently large,*

$$\bar{d}\left(\{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}, \{Y_i(x)\}_{(x,i) \in B_{a,b}}\right) \le 8\varepsilon^{1/4}$$

for all $\eta \notin E$, where $B_{a,b}$ is defined as before.

In Proposition 5.5, $\tilde{\nu}^\eta$ denotes the measure on $X^{\mathbf{N}}$ induced by the random variables $\{Y_i^\eta\}_{i \ge 0}$.

PROPOSITION 5.5. *If the spin rates are finite range, then given any positive integer b and $\varepsilon > 0$, for all a sufficiently large, the following holds: If $\eta \in X$ and $\{l_1, \dots, l_m\} \subseteq (B_a)^c \times [1, b]$, then for all atoms A in $\mathscr{A}(\sigma(l_1, \dots, l_m))$ except for ε portion (with respect to $\tilde{\nu}^\eta$),*

$$\bar{d}\left(\{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}, \{Y_i^\eta(x)\}_{(x,i) \in B_{a,b}}/A\right) \le \varepsilon.$$

We call an A satisfying this last inequality η-good, although there is also an obvious dependence on a and b. We now proceed with the proof of Theorem 5.3. Combining Propositions 5.4 and 5.5, we obtain the following.

Given $\varepsilon > 0$, there is a set $E \subseteq X$ of $\underline{\nu}$-measure less than ε and a box $B_{a,b} \subseteq \mathbf{Z}^{n+1}$ such that for all $\eta \notin E$ and for all $\{l_1, \dots, l_m\} \subseteq (B_a)^c \times [1, b]$,

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i)\in B_{a,b}}/A, \{Y_i(x)\}_{(x,i)\in B_{a,b}}\big) \le 9\varepsilon^{1/4}$$

for all atoms A in $\mathscr{A}(\sigma(l_1, \dots, l_m))$ except for ε portion (with respect to $\tilde{\nu}^\eta$).

We now proceed analogously to Theorem 4.6. Let

$$\{y_1, \dots, y_r\} \subseteq (B_{a,b})^c \cap \{\omega \in \mathbf{Z}^{n+1}: \omega_{n+1} \le b\}.$$

Partition $\{y_1, \dots, y_r\}$ into two sets $\{l_1, \dots, l_m\}$ and $\{l'_1, \dots, l'_{m'}\}$, where $l_1, \dots, l_m$ all have their time coordinate positive and $l'_1, \dots, l'_{m'}$ all have their time coordinate nonpositive. Then any element in $\mathscr{A}(\sigma(y_1, \dots, y_r))$ is $A_1 \cap A_2$, where $A_1 \in \mathscr{A}(\sigma(l_1, \dots, l_m))$ and $A_2 \in \mathscr{A}(\sigma(l'_1, \dots, l'_{m'}))$.

Again let Q^{A_2} denote the measure on X by considering the conditional distribution of Y_0 given A_2. We call A_2 a good atom if $Q^{A_2}(E) \le \sqrt{\varepsilon}$. Since $\underline{\nu}(E) < \varepsilon$, all but $\sqrt{\varepsilon}$ portion (with respect to $\tilde{\underline{\nu}}$) of the atoms A_2 are good atoms. We call A_1 an A_2-good atom if

$$Q^{A_2}\{\eta: A_1 \text{ is } \eta\text{-good}\} \ge 1 - \sqrt{\varepsilon}.$$

Proposition 5.5 implies that for any A_2 atom, at least $1 - \sqrt{\varepsilon}$ portion [with respect to $\int_X \tilde{\nu}^\eta \, dQ^{A_2}(\eta)$] of the atoms A_1 are A_2-good. Note that $\int_X \tilde{\nu}^\eta \, dQ^{A_2}(\eta)$ is just the measure on $X^{\mathbf{Z}}$ obtained by conditioning on the atom A_2. It follows that all but $2\sqrt{\varepsilon}$ (with respect to $\tilde{\underline{\nu}}$) of the atoms $A = A_1 \cap A_2$ in $\mathscr{A}(\sigma(y_1, \dots, y_r))$ are such that A_2 is good and A_1 is A_2-good.

To complete the proof of the VWB property, we show that if A_2 is a good atom and A_1 is an A_2-good atom, then

$$\bar{d}\big(\{Y_i(x)\}_{(x,i)\in B_{a,b}}/(A_1 \cap A_2), \{Y_i(x)\}_{(x,i)\in B_{a,b}}\big) \le 9\varepsilon^{1/4} + 2\sqrt{\varepsilon}.$$

If $\eta \notin E$ and A_1 is η-good, let P^{η, A_1} be a

$$\{Y_i^\eta(x)\}_{(x,i)\in B_{a,b}}/A_1 - \{Y_i(x)\}_{(x,i)\in B_{a,b}}$$

coupling which gives them $\bar{d}$-distance less than or equal to $9\varepsilon^{1/4}$. If $\eta \in E$ or $\eta \notin E$ and A_1 is not η-good, let P^{η, A_1} be an arbitrary coupling of the previous two processes. Then $P^{A_1, A_2} = \int_X P^{\eta, A_1} \, dQ^{A_2}(\eta)$ is a

$$\{Y_i(x)\}_{(x,i)\in B_{a,b}}/(A_1 \cap A_2) - \{Y_i(x)\}_{(x,i)\in B_{a,b}}$$

coupling satisfying

$$\begin{aligned}
&\bar{d}_{P^{A_1, A_2}}\big(\{Y_i(x)\}_{(x,i)\in B_{a,b}}/(A_1 \cap A_2), \{Y_i(x)\}_{(x,i)\in B_{a,b}}\big) \\
&\quad = \int_X \bar{d}_{P^{\eta, A_1}}\big(\{Y_i^\eta(x)\}_{(x,i)\in B_{a,b}}/A_1, \{Y_i(x)\}_{(x,i)\in B_{a,b}}\big) \, dQ^{A_2}(\eta) \\
&\quad \le 9\varepsilon^{1/4} + \sqrt{\varepsilon} + \sqrt{\varepsilon}
\end{aligned}$$

by integrating over the two pieces

$$(E)^c \cap \{\eta: A_1 \text{ is } \eta\text{-good}\}$$

and its complement. This demonstrates the VWB property. □

We now prove Proposition 5.5.

PROOF OF PROPOSITION 5.5. In proving this proposition, the problem is to recognize when the atoms in $\mathscr{A}(\sigma(l_1, \dots, l_m))$ affect the evolution in $B_{a,b}$. We therefore couple the process with a modified process where there is no effect of the atoms of $\mathscr{A}(\sigma(l_1, \dots, l_m))$ on $B_{a,b}$ and show that this coupling is good in a certain sense.

If D is a box in $\mathbf{Z}^n$ of the form $\prod_{i=1}^n[-c, c]$, let T^D denote the Markov operator corresponding to the Markov kernel obtained from modifying the spin rates so that points inside and outside of D evolve independently. These spin rates should therefore be

$$c_D(x, \delta) = c(x, {}^x\delta^D),$$

where

$${}^x\delta^D = \begin{cases} \delta(y) & \text{if } x, y \in D \text{ or } x, y \in D^c, \\ 0 & \text{if } x \in D, y \in D^c \text{ or } x \in D^c, y \in D. \end{cases}$$

We let $P^{\eta, D}$ denote the resulting measure on $(X \times X)^{\mathbf{N}}$ corresponding to the basic coupling of $\{c(x, \delta)\}$ and $\{c_D(x, \delta)\}$ with initial distribution $\delta_\eta \times \delta_\eta$, and we let Y^1 and Y^2 be the two marginal processes. Next, if B is a box in $\mathbf{Z}^n$ of the form $\prod_{i=1}^n[-a, a]$, we let B^ε denote the largest box containing B of this same form and satisfying $|B^\varepsilon| \le (1 + \varepsilon)|B|$.

Ideas from first-passage percolation theory allow us to establish Lemma 5.6 which enables us to complete the proof of Proposition 5.5. The proof of this lemma is given afterward.

LEMMA 5.6. *If the spin rates are finite range, then for all $\varepsilon > 0$ and $b \in \mathbf{N}$, for all sufficiently large boxes $B \subseteq \mathbf{Z}^n$ of the form $\prod_{i=1}^n[-a, a]$,*

$$P^{\eta, B^\varepsilon}\Big\{Y_i^1(x) \ne Y_i^2(x) \text{ for some } x \in B \cup (B^{2\varepsilon})^c \text{ and } i \in [1, b]\Big\} \le \varepsilon$$

for all $\eta \in X$.

By Lemma 5.6 choose B sufficiently large so the conclusion of this lemma holds. Now let a be chosen so that $\prod_{i=1}^n[-a, a] = B^{2\varepsilon}$. We claim that the conclusion of Proposition 5.5 holds with this particular a.

Consider the projection of P^{η, B^ε} onto $(B \times [1, b]) \cup \{l_1, \dots, l_m\}$. The two marginals of this factor corresponding to Y^1 and Y^2 give us two measures μ_1 and μ_2 on

$$\{0, 1\}^{(B \times [1, b]) \cup \{l_1, \dots, l_m\}}$$

satisfying

$$\sum_{z} |\mu_1(z) - \mu_2(z)| \le 2\varepsilon \tag{5.1}$$

by Lemma 5.6 where the sum extends over all $z \in \{0,1\}^{(B\times[1,b])\cup\{l_1,\dots,l_m\}}$. Let $\mathscr{Q} = \{Q_i\}$ denote the partition of

$$\{0,1\}^{(B\times[1,b])\cup\{l_1,\dots,l_m\}}$$

into $2^{|B\times[1,b]|}$ sets according to $B\times[1,b]$ and let $\mathscr{R} = \{R_i\}$ denote the analogous partition into 2^m sets corresponding to $\{l_1,\dots,l_m\}$. Clearly $\mathscr{Q}$ and $\mathscr{R}$ are μ_2-independent. This together with (5.1) implies that

$$\sum_{i,j} |\mu_1(Q_i\cap R_j) - \mu_1(Q_i)\mu_1(R_j)| \le 6\varepsilon,$$

since

$$\begin{aligned}
&\sum_{i,j} |\mu_1(Q_i\cap R_j) - \mu_1(Q_i)\mu_1(R_j)| \\
&\quad\le \sum_{i,j} |\mu_1(Q_i\cap R_j) - \mu_2(Q_i\cap R_j)| + \sum_{i,j} |\mu_2(Q_i)\mu_2(R_j) - \mu_1(Q_i)\mu_1(R_j)| \\
&\qquad [\text{since } \mathscr{Q} \text{ and } \mathscr{R} \text{ are } \mu_2\text{-independent}] \\
&\quad\le 2\varepsilon + \sum_{i,j} |\mu_2(Q_i)\mu_2(R_j) - \mu_2(Q_i)\mu_1(R_j)| \\
&\qquad + \sum_{i,j} |\mu_2(Q_i)\mu_1(R_j) - \mu_1(Q_i)\mu_1(R_j)| \le 6\varepsilon.
\end{aligned}$$

From this it follows [see Shields (1973)] that for all but $\sqrt{6\varepsilon}$ portion of the atoms R_i in $\mathscr{R}$ (with respect to μ_1),

$$\sum_{j} |\mu_1(Q_j) - \mu_1(Q_j/R_i)| \le \sqrt{6\varepsilon},$$

which implies

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i)\in B\times[1,b]}, \{Y_i^\eta(x)\}_{(x,i)\in B\times[1,b]}/R_i\big) \le \sqrt{6\varepsilon}.$$

Since $|B^{2\varepsilon}| \le (1+2\varepsilon)|B|$, it then follows that

$$\bar{d}\big(\{Y_i^\eta(x)\}_{(x,i)\in B^{2\varepsilon}\times[1,b]}, \{Y_i^\eta(x)\}_{(x,i)\in B^{2\varepsilon}\times[1,b]}/R_i\big) \le \sqrt{6\varepsilon} + 2\varepsilon.$$

This proves the proposition with ε replaced by $\sqrt{6\varepsilon} + 2\varepsilon$. □

PROOF OF LEMMA 5.6. We prove the stronger conclusion that

$$E^{P^{\eta,B^\varepsilon}}\left[\sum_{\substack{x\in B\cup(B^{2\varepsilon})^c\\ i\in[1,b]}} I_{\{Y_i^1(x)\ne Y_i^2(x)\}}\right] \le \varepsilon \tag{5.2}$$

for all $\eta \in X$ if B is sufficiently large. Since the spin rates have finite range, there is some integer F such that $e_x(0) = 0$ for all x satisfying $|x| > F$. Since all the spin rates are bounded by 1 (in fact $\frac{1}{2}$), we reduce this to another problem. We construct the following auxiliary IPS:

$$\tilde{c}(0,\delta) = \begin{cases} 0 & \text{if } \delta(0) = 1, \\ 0 & \text{if } \delta(0) = 0 \text{ and } \delta(y) = 0 \text{ for all } y \text{ with } |y| \le F, \\ k & \text{if } \delta(0) = 0 \text{ and } |\{y\colon \delta(y) = 1 \text{ and } |y| \le F\}| = k. \end{cases}$$

Assuming translation invariance, we have defined $\tilde{c}(x,\delta)$ for all x and δ. If 1 is interpreted as sick and 0 as healthy, this system is a model for the spread of disease where no one is ever cured and where a healthy individual becomes sick at rate k when there are k sick individuals within F units.

We shall show that all sufficiently large boxes B satisfy

$$E^{\partial B^\varepsilon}\left[\# \text{ of 1's in } B \cup (B^{2\varepsilon})^c \text{ at time } b\right] \le \varepsilon, \tag{5.3}$$

where the rates $\{\tilde{c}(x,\eta)\}$ are being used and where the superscript ∂B^ε means that one is taking the initial configuration to be 1's on the boundary of B^ε and 0's elsewhere. Since the collection of lattice points where the Y^1 process does not equal the Y^2 process can be dominated by the lattice points which are in state 1 in the auxiliary system, it follows that (5.3) implies (5.2).

Let N denote the length of one side of the box B^ε. Simple geometry then implies the existence of a constant C_n (depending only on the dimension n of the lattice) so that

$$d^1(\partial B, \partial B^\varepsilon) \ge C_n \varepsilon N$$

and

$$d^1(\partial B^\varepsilon, \partial B^{2\varepsilon}) \ge C_n \varepsilon N.$$

Here d^1 denotes the L^1 distance on $\mathbf{Z}^n$ between two sets. A sequence γ of distinct lattice points $(x_0,\dots,x_l)$ is called an F-path if $|x_i - x_{i+1}| \le F$ for each i. We let $|\gamma|$ denote the number of elements in the F-path γ.

A more convenient way to view this problem is using the *graphical representation* as discussed in Liggett (1985). In our case, this becomes the following: First, for every $x_i \ne x_j$ satisfying $|x_i - x_j| \le F$, we let $Z(x_i, x_j)$ be an exponentially distributed random variable with parameter 1. We take the family of random variables $\{Z(x_i,x_j)\}$ to be independent. Now we construct a new process: At time t, lattice point x is 1 if and only if there is an F-path $\gamma = (x_0,\dots,x_l)$ from ∂B^ε to x satisfying $\sum_{i=0}^{l-1} Z(x_i,x_{i+1}) \le t$. This can easily be seen to be the same system we defined above with initial configuration ∂B^ε. Moreover, it is easier to make calculations using this graphical representation.

Note that any F-path from ∂B^ε to $B \cup (B^{2\varepsilon})^c$ must have $|\gamma| \ge C_n \varepsilon N/F$. If $\gamma = (x_0,\dots,x_l)$, we call $\sum_{i=0}^{l-1} Z(x_i,x_{i+1})$ the time of γ and denote it by T_γ. This is, of course, a random variable. Introducing some convenient notation, we let

$$\Gamma_x = \{\gamma\colon \gamma \text{ is an } F\text{-path from } \partial B^\varepsilon \text{ to } x\},$$

$$\Gamma = \left\{\gamma\colon \gamma \text{ is an } F\text{-path from } \partial B^\varepsilon \text{ to } B \cup (B^{2\varepsilon})^c\right\},$$

$$\Gamma_j = \left\{\gamma\colon \gamma \text{ is an } F\text{-path from } \partial B^\varepsilon \text{ to } B \cup (B^{2\varepsilon})^c \text{ with } |\gamma| = j\right\}.$$

In addition to these, we let $K(n, F)$ be the number of lattice points in the closed ball of radius F contained in $\mathbf{Z}^n$ in the L^1 metric and let $P(j, b)$ be the probability that a Poisson process with parameter 1 has j occurrences by time b. Finally, in the graphical representation,

$$
\begin{aligned}
&E^{\partial B^\varepsilon}\Big[\#\text{ of 1's in } B \cup (B^{2\varepsilon})^c \text{ at time } b\Big] \\
&\quad= \sum_{x \in B \cup (B^{2\varepsilon})^c} P^{\partial B^\varepsilon}(\text{there is a 1 at } x \text{ at time } b) \\
&\quad\le \sum_{x \in B \cup (B^{2\varepsilon})^c} \sum_{\gamma \in \Gamma_x} P^{\partial B^\varepsilon}\{T_\gamma \le b\} \\
&\quad= \sum_{\gamma \in \Gamma} P^{\partial B^\varepsilon}\{T_\gamma \le b\} \\
&\quad= \sum_{j \ge \lceil C_n \varepsilon N/F \rceil} \sum_{\gamma \in \Gamma_j} P^{\partial B^\varepsilon}\{T_\gamma \le b\} \\
&\quad\le \sum_{j \ge \lceil C_n \varepsilon N/F \rceil} 2nN^{n-1} K(n, F)^j P(j, b) \\
&\quad= \sum_{j \ge \lceil C_n \varepsilon N/F \rceil} 2nN^{n-1} K(n, F)^j \sum_{r \ge j} \frac{e^{-b} b^r}{r!} \\
&\quad\le \sum_{j \ge \lceil C_n \varepsilon N/F \rceil} 2nN^{n-1} K(n, F)^j C(n, F, b) \frac{1}{(K(n, F) + 1)^j} \\
&\qquad\qquad [\text{where } C(n, F, b) \text{ is some constant depending only on } n,\ F \text{ and } b] \\
&\quad= 2nN^{n-1} C(n, F, b) \sum_{j \ge \lceil C_n \varepsilon N/F \rceil} \left(\frac{K(n, F)}{K(n, F) + 1} \right)^j \quad \text{which} \to 0 \text{ as } N \to \infty.
\end{aligned}
$$

□

THEOREM 5.7. *If the spin rates are attractive, then* $(D_X(\mathbf{R}), \mathscr{B}, \underline{\tilde{\nu}}, \mathbf{R} \times \mathbf{Z}^n)$ *is Bernoulli where* $D_X(\mathbf{R})$ *and* $\mathscr{B}$ *are as defined above and where* $\mathbf{R} \times \mathbf{Z}^n$ *acts canonically on* $D_X(\mathbf{R})$.

PROOF. The proof of Theorem 4.7 using Theorem 5.3 instead of Theorem 4.6 can be repeated here almost verbatim replacing sums by appropriate integrals. □

6. Extensions. Possible extensions would be to replace the one particle state space $\{0, 1\}$ by a more general partially ordered space, consider a different type of dynamics or replace $\mathbf{Z}^n$ by a more general group. From a pure ergodic theoretic point of view, the more general group should at least be amenable since these are the groups to which classical ergodic theory can be extended. In another paper Steif (1990), we prove all of the previous results plus more for another general family of spin systems, not necessarily attractive.

Acknowledgment. Finally, I would like to express my appreciation for the help and encouragement given to me by my thesis advisor, Professor Donald S. Ornstein.

REFERENCES

CONZE, J. P. (1972). Entropie d'un group abelien de transformations. *Z. Wahrsch. Verw. Gebiete* **25** 11–30.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes—Characterization and Convergence*. Wiley, New York.

FELDMAN, J. (1980). r-Entropy, equipartition, and Ornstein's isomorphism theorem in $\mathbf{R}^n$. *Israel J. Math.* **36** 321–343.

KATZNELSON, Y. and WEISS, B. (1972). Commuting measure-preserving transformations. *Israel. J. Math.* **12** 161–173.

LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.

ORNSTEIN, D. S. (1974). *Ergodic Theory, Randomness and Dynamical Systems*. Yale Univ. Press, New Haven, Conn.

ORNSTEIN, D. S. and WEISS, B. (1974). Finitely determined implies very weak Bernoulli. *Israel J. Math.* **17** 94–104.

ORSTEIN, D. S. and WEISS, B. (1987). Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.* **48** 1–141.

ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. Springer, New York.

SHIELDS, P. (1973). *The Theory of Bernoulli Shifts*. Univ. Chicago Press.

STEIF, J. E. (1990). $\bar{d}$-convergence to equilibrium and space-time Bernoullicity for spin systems in the $M < \varepsilon$ case. Unpublished manuscript.

THOUVENOT, J. P. (1972). Convergence en moyenne de l'information pour l'action de $\mathbf{Z}^2$. *Z. Wahrsch. Verw. Gebiete* **24** 135–137.

WALTERS, P. (1975). *An Introduction to Ergodic Theory*. Springer, New York.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853