Mixing in Time and Space for Lattice Spin Systems: A Combinatorial View

Martin Dyer [*] Alistair Sinclair [†] Eric Vigoda [‡] Dror Weitz [§]

November 5, 2003

Abstract

The paper considers spin systems on the d-dimensional integer lattice $\mathbb{Z}^d$ with nearest-neighbor interactions. A sharp equivalence is proved between decay with distance of spin correlations (a *spatial* property of the equilibrium state) and rapid mixing of the Glauber dynamics (a *temporal* property of a Markov chain Monte Carlo algorithm). Specifically, we show that if the mixing time of the Glauber dynamics is $O(n \log n)$ then spin correlations decay exponentially fast with distance. We also prove the converse implication for monotone systems, and for general systems we prove that exponential decay of correlations implies $O(n \log n)$ mixing time of a dynamics that updates sufficiently large blocks (rather than single sites). While the above equivalence was already known to hold in various forms, we give proofs that are purely combinatorial and avoid the functional analysis machinery employed in previous proofs.

[*] School of Computing, University of Leeds, Leeds LS2 9JT, United Kingdom. (dyer@comp.leeds.ac.uk). Supported by EPSRC grant "Sharper Analysis of Randomised Algorithms: a Computational Approach" and by EC IST project RAND-APX.

[†] Computer Science Division, University of California, Berkeley, Berkeley, CA 94720, U.S.A. (sinclair@cs.berkeley.edu). Supported in part by NSF grants CCR-9820951 and CCR-0121555, and by DARPA cooperative agreement F30602-00-2-060.

[‡] Department of Computer Science, University of Chicago, Chicago, IL 60637, U.S.A. (vigoda@cs.uchicago.edu)

[§] Computer Science Division, University of California, Berkeley, Berkeley, CA 94720, U.S.A. (dror@cs.berkeley.edu). Supported in part by NSF grants CCR-9820951 and CCR-0121555, and by DARPA cooperative agreement F30602-00-2-060.

1 Introduction

Lattice spin systems are a class of models that originated in Statistical Physics, though interest in them has since expanded to many other areas, including Probability Theory, Statistics, Artificial Intelligence, and Theoretical Computer Science. A *(lattice) spin system* consists of a collection of sites which are the vertices of a regular lattice graph. A *configuration* of the spin system is an assignment of one of a finite set of *spins* to each site. The sites interact locally, according to potentials specified by the system, such that different combinations of spins on neighboring sites have different relative likelihoods. This interaction gives rise to a well-defined probability distribution over configurations of any finite subset (volume) of the sites, conditional on a fixed configuration of the sites on the boundary of this subset. Such a distribution is referred to as a *finite volume Gibbs distribution*, and is regarded as the equilibrium state of the given subset conditional on the given boundary configuration.

A *Glauber dynamics* is a Markov chain Monte Carlo algorithm used to sample from the Gibbs distribution. A step in this Markov chain is a random update of the spin of a single site (or of a finite set of sites), conditional on its neighboring spins and in a manner which is reversible with respect to the Gibbs distribution. As a result, such a Markov chain converges to the corresponding Gibbs distribution. The Glauber dynamics plays a central role not just as an algorithm for sampling from the Gibbs distribution but also as a model for the physical process of reaching equilibrium.

A striking phenomenon in the field of spin systems, at least for lattices with "sub-exponential growth" such as the integer lattice $\mathbb{Z}^d$, is the equivalence of (*a priori* unrelated) notions of temporal and spatial mixing. By *temporal mixing* we mean that the Glauber dynamics converges "very fast" to its stationary Gibbs distribution, while by *spatial mixing* we mean that in the Gibbs distribution, correlations between the spins of different sites decay "very fast" with the (lattice) distance between them. This equivalence is interesting because it precisely relates the running time of an algorithm to purely spatial properties of the underlying model. In addition, a common heuristic in computer science is that local algorithms should work well (run fast) for local problems. The equivalence between temporal and spatial mixing is an example of the above heuristic in a restricted setting, where the relationship is formally proven and where there are precise interpretations for the terms "local algorithm", "local problem", and "run fast".

The above equivalence has been explored by a number of previous authors, using various notions of spatial and temporal mixing. This line of work was initiated by Holley [10] and Aizenman and Holley [1], followed by Zegarlinski [18] and culminating in the work of Stroock and Zegarlinski [17], who were the first to establish the above equivalence in full. We further mention Martinelli and Olivieri [13, 14], who later obtained sharper results by working with a weaker spatial mixing assumption, and Cesi [4], who recently simplified some of the proofs. See also [12] for a review of results in the field.

The references mentioned above make crucial use of functional analysis in their proofs, and usually discuss quantities such as the *spectral gap* and the *logarithmic Sobolev constant* of the dynamics as a measure of its temporal mixing (these quantities measure the contraction of the semi-group associated with the dynamics). In this paper, we give purely combinatorial proofs of this equivalence, based on the elementary technique of coupling probability distributions. Although some of the ideas we use have appeared before, our main contribution lies in presenting a complete argument which is purely combinatorial, where the reader does not need to resort to concepts from functional analysis.

We note that the result we present in the direction going from spatial mixing to temporal mixing (of the single-site Glauber dynamics) is limited in the sense that it only applies to *monotone* systems. For general systems, however, we show that spatial mixing implies temporal mixing of a "finite-block" Glauber dynamics, in which a sufficiently large block of spins is updated at each step. The corresponding implication for the single-site dynamics in the general case is known [4, 12, 14, 17], but currently we do not have a combinatorial proof of it.

The remainder of the paper is organized as follows. Section 2 includes exact definitions and statements of results. In Sect. 3 we list a few basic tools we use in the proofs. In Sect. 4 we prove that temporal mixing implies spatial mixing while in Sects. 5 and 6 we prove that spatial mixing implies temporal mixing for monotone and general systems respectively.

2 Definitions and statements of results

2.1 Spin systems

Consider the d-dimensional integer lattice[1] as a graph with vertex set $V = \mathbb{Z}^d$ and edge set E, where $(v,u) \in E$, denoted $v \sim u$, if and only if $\sum_{i=1}^{d} |v_i - u_i| = 1$. We use the statistical physics terminology and refer to the vertices as *sites*. For a finite subset $\Psi \subset V$, we define its *boundary* as

$$\partial\Psi = \{v \notin \Psi : \text{ there exists } u \in \Psi \text{ s.t. } v \sim u\}.$$

Each site is assigned a spin from the spin space $S = \{1, \ldots, q\}$, and the configuration space is denoted by $\Omega = \Omega_\Psi = S^\Psi$. Given a configuration $\sigma \in \Omega$, we write $\sigma[v]$ for the spin that σ assigns to v and abuse this notation with $\sigma[\Lambda]$ standing for the configuration of the subset Λ under σ.

We consider spin systems with nearest neighbor interactions (although everything we do can be generalized to finite range interactions). Namely, we have a (symmetric) pair potential[2] $U : S \times S \to \mathbb{R}$, and a self potential $W : S \to \mathbb{R}$. Then, for a finite subset Ψ and a boundary configuration $\tau \in \Omega_{\partial\Psi}$, the Hamiltonian $H_\Psi^\tau : \Omega_\Psi \to \mathbb{R}$ is defined as

$$H_\Psi^\tau(\sigma) \;=\; \sum_{v \in \partial\Psi, u \in \Psi, v \sim u} U(\tau[v], \sigma[u]) \;+\; \sum_{v,u \in \Psi, v \sim u} U(\sigma[v], \sigma[u]) \;+\; \sum_{v \in \Psi} W(\sigma[v]).$$

The value this Hamiltonian assigns can be considered as the "energy" of σ when τ is the boundary configuration. The *finite volume Gibbs distribution* associated with the subset Ψ and the boundary configuration τ assigns probability to σ which is proportional to the inverse exponential of its energy. Formally,

$$\mu_\Psi^\tau(\sigma) = \frac{1}{Z_\Psi^\tau} \exp(-H_\Psi^\tau(\sigma)), \tag{1}$$

where Z_Ψ^τ is the appropriate normalizing factor.

Example: Probably the best known spin system is the *ferromagnetic Ising model*. In this case, the spin space is $S = \{-1, +1\}$, while $U(s_1, s_2) = -\beta \cdot s_1 \cdot s_2$ and $W(s) = -\beta \cdot h \cdot s$, where $\beta \in \mathbb{R}^+$ is

[1] Most of our results hold — with suitable modifications — for any lattice with "sub-exponential growth" (i.e., the volume of increasing balls around any site increases sub-exponentially with the radius). For simplicity, in this paper we focus just on $\mathbb{Z}^d$.

[2] The given definition of the pair potential does not cover systems with hard constraints, where U may be infinite. Systems with hard constraints are discussed in section 2.5 below.

the inverse temperature and $h \in \mathbb{R}$ is an external field. Thus, the energy of a configuration is linear in the number of edges with disagreeing spins, as well as the number of spins with sign opposite to that of h. For example, if $h = 0$ and if we ignore the effect of the boundary configuration (the so-called "free-boundary condition") then the minimum energy (highest probability) configurations are the two constant configurations where all the spins have the same value (either $+1$ or -1).

2.2 The Glauber dynamics

We study the following simple Markov chain (X_t), known as the (heat-bath) *Glauber dynamics*, which is used to sample from μ_Ψ^τ. Given the current configuration $X_t \in \Omega_\Psi$, the transition $X_t \to X_{t+1}$ is defined as follows:

- Choose a vertex v uniformly at random from Ψ.
- Let $X_{t+1}[u] = X_t[u]$ for all $u \neq v$.
- Choose $X_{t+1}[v]$ from $\mu_{\{v\}}^{X'_t}$, where X'_t is the configuration of $\partial\{v\}$ defined by $X'_t[u] = X_t[u]$ for $u \in \Psi$ and $X'_t[u] = \tau[u]$ for $u \in \partial\Psi$.

It is not too difficult to verify that this Markov chain is reversible with respect to the Gibbs distribution μ_Ψ^τ and, in particular, that μ_Ψ^τ is the unique stationary distribution.

Remark: In the literature, a Glauber dynamics is usually any Markov chain that makes single-site updates that are reversible with respect to the single-site Gibbs measure. Indeed, all the results below apply to any choice of Glauber dynamics. However, for definiteness we will assume the above definition throughout this paper.

We also discuss a generalization of the Glauber dynamics to a Markov chain where at each step a block of sites is updated rather than a single site. Let $Q_L = [1, \ldots, L]^d$ be the d-dimensional regular box of side length L. Consider all the translations of Q_L that intersect the subset Ψ and let $B(\Psi, L) = \{\Lambda \neq \emptyset \mid \Lambda = (z + Q_L) \cap \Psi \text{ for some } z \in \mathbb{Z}^d\}$. We think of each $\Lambda \in B(\Psi, L)$ as a *block*. We then denote by $HB(L)$ the heat-bath block dynamics that makes updates to blocks from $B(\Psi, L)$. Given the current configuration X_t, the transition $X_t \to X_{t+1}$ is defined as follows:

- Choose a block Λ uniformly at random from $B(\Psi, L)$.
- Let $X_{t+1}[u] = X_t[u]$ for all $u \notin \Lambda$.
- Choose $X_{t+1}[\Lambda]$ from $\mu_\Lambda^{X'_t}$, where X'_t is the configuration of $\partial\Lambda$ defined by $X'_t[u] = X_t[u]$ for $u \in \Psi$ and $X'_t[u] = \tau[u]$ for $u \in \partial\Psi$.

2.3 Temporal and Spatial Mixing

The statements in this paper relate an appropriate notion of *temporal mixing* (convergence in time of the Glauber dynamics) with an appropriate notion of *spatial mixing* (decay of correlation with distance in the Gibbs distribution). The exact definitions are given below.

Let μ_1 and μ_2 be two distributions on Ω_Ψ. We write $\|\mu_1 - \mu_2\| = \max_{A \subseteq \Omega_\Psi} |\mu_1(A) - \mu_2(A)|$ for the *total variation distance* between the two distributions, and $\|\mu_1 - \mu_2\|_\Lambda = \max_{A \subseteq \Omega_\Lambda} |\mu_1(A) - \mu_2(A)|$ for the distance when projecting the two distributions on Ω_Λ for $\Lambda \subseteq \Psi$.

Definition 2.1. *We say that the Glauber dynamics has* optimal temporal mixing *if there exist constants b and $c > 0$ such that for any subset Ψ with any boundary configuration, the dynamics on Ψ has the following property. For any two instances (X_t) and (Y_t) of the chain and for any positive integer k, $\|X_{kn} - Y_{kn}\| \leq bn\exp(-ck)$, where $n = |\Psi|$ is the volume of Ψ.*

In particular, optimal temporal mixing means that the distance from the stationary measure $\|X_{kn} - \mu_\Lambda^\tau\| \leq bn\exp(-ck)$ for any instance (X_t). Before we move on to the definition of the spatial mixing notion, we pause to compare optimal temporal mixing as defined here with some of the other notions of temporal mixing found in the literature. The *mixing time* of a Markov chain (as a function of ϵ) is the time it takes to get within a variation distance of ϵ from the stationary measure. Notice that optimal temporal mixing is equivalent to a mixing time of $O(n\log(\frac{n}{\epsilon}))$. Optimal temporal mixing also implies that the spectral gap of the dynamics is at least $\frac{c}{n}$. While such a spectral gap does not immediately imply optimal temporal mixing, it is not too difficult to see that if the log-Sobolev constant associated with the dynamics is bounded from below by $\frac{c}{n}$ then the dynamics has optimal temporal mixing. We notice that in fact, in the context of spin systems, all the above notions of temporal mixing are known to be equivalent when considered to hold uniformly in the subset Ψ and in the boundary configuration (since they are all equivalent to an appropriate notion of spatial mixing as below).

Remark: We note that the word *optimal* in Definition 2.1 should not be taken literally. Although it is indeed believed that the mixing time of the Glauber dynamics cannot be $o(n\log n)$, there is not yet a rigorous proof of this conjecture for general spin systems.

The corresponding spatial notion we consider states that changing the spin of a site on the boundary has an exponentially small effect on the configuration of sites far away from the changed site. The distance between two sites v and u is defined as the graph distance between them, or equivalently, $\mathrm{dist}(v,u) = \sum_{i=1}^d |v_i - u_i|$. The distance between subsets is the natural extension, i.e., the minimal distance between two sites, one in each subset.

Definition 2.2. *We say the system has* strong spatial mixing *if there exist constants β and $\alpha > 0$ such that for any two subsets Λ, Ψ where $\Lambda \subseteq \Psi$, any site $u \in \partial\Psi$, and any pair of boundary configurations τ and τ^u that differ only at u, $\|\mu_\Psi^\tau - \mu_\Psi^{\tau^u}\|_\Lambda \leq \beta|\Lambda|\exp(-\alpha\cdot\mathrm{dist}(u,\Lambda))$.*

Remark: In the literature, the definition of strong spatial mixing may vary, where the difference lies in which class of subsets Ψ the assumption applies to (for example, Ψ may be restricted to be a regular box). We work with the strongest version by requiring it to apply to all subsets in order to simplify our arguments.

In order to illustrate the above definitions[3], let us conclude this section with a brief discussion of how they apply to the Ising model (as defined in Example 2.1) on the square lattice $\mathbb{Z}^2$. Recall that in the definition of the Ising model, β stands for the inverse temperature and h for an external field. The following fact is an example of the equivalence between temporal and spatial mixing: There exists a critical $\hat{\beta}_c$ such that, when $h = 0$ (no external field), for $\beta < \hat{\beta}_c$ both optimal temporal

[3]Strictly speaking, the discussion in the three paragraphs starting here applies to slightly modified definitions of spatial and temporal mixing where the subset Ψ is restricted to have a "nice" shape (see remark following Definition 2.2).

mixing (Definition 2.1) and strong spatial mixing (Definition 2.2) hold for the Ising model on $\mathbb{Z}^2$, while for $\beta > \hat{\beta}_c$ both fail.

It is worth mentioning here that in the special case of the Ising model on $\mathbb{Z}^2$, the critical $\hat{\beta}_c$ mentioned above coincides [15] with the critical inverse temperature β_c where a phase transition occurs in the infinite volume limit, namely, for $\beta < \beta_c$ there exists a unique infinite volume Gibbs measure while for $\beta > \beta_c$ there are multiple such measures. Though we do not discuss infinite volume Gibbs measures in this paper (see for example [8, 9] for more on this topic), one can interpret the uniqueness of the infinite volume Gibbs measure as an alternative notion of spatial mixing (which is weaker than strong spatial mixing provided the underlying lattice is of sub-exponential growth). Notice that in general it is not true that the two critical inverse temperatures $\hat{\beta}_c$ and β_c coincide, and there are examples where the infinite volume Gibbs measure is unique while strong spatial mixing does not hold (see [12] for a discussion on the matter).

Finally, again in the special case of the Ising model on $\mathbb{Z}^2$, the corresponding phase transition in the mixing time is known to be very sharp [5]. Specifically, for $\beta > \hat{\beta}_c = \beta_c$, not only does optimal temporal mixing not hold, but in fact the mixing time is super-polynomial (specifically, $\exp(c\sqrt{n})$ for some constant $c > 0$).

2.4 Monotone systems

Some of the statements in this paper apply only to monotone systems. In a monotone system, each site v is associated with a linear ordering of the spin space, denoted by $\succeq_v$. Since the spin space is finite, each of the linear orderings has unique maximal and minimal elements, which we call the *plus* and *minus* elements respectively. The single-site orderings give rise to a partial ordering $\succeq_\Psi$ of the configuration space. Specifically, $\sigma_1 \succeq_\Psi \sigma_2$ if and only if $\sigma_1[v] \succeq_v \sigma_2[v]$ for every $v \in \Psi$. The system is *monotone* with respect to the above partial ordering if, for every subset Ψ and any two boundary configurations τ_1 and τ_2 such that $\tau_1 \succeq_{\partial\Psi} \tau_2$, the Gibbs measure $\mu_\Psi^{\tau_1}$ statistically dominates the Gibbs measure $\mu_\Psi^{\tau_2}$ with respect to $\succeq_\Psi$. Equivalently, the two distributions can be coupled such that with probability 1, $\sigma_1 \succeq_\Psi \sigma_2$, where σ_1 and σ_2 are a pair of coupled configurations chosen from $\mu_\Psi^{\tau_1}$ and $\mu_\Psi^{\tau_2}$ respectively. Notice that it is enough that the above property holds for all single sites to ensure that it holds for all subsets Ψ. Also, since the single-site orderings are linear, the system is "realizably" monotone [7]. This means that, given a collection of boundary configurations $\tau_1, \tau_2, \ldots, \tau_k$, we can simultaneously couple the k corresponding Gibbs distributions such that if $\tau_i \succeq_{\partial\Psi} \tau_j$, the corresponding coupled configurations satisfy $\sigma_i \succeq_\Psi \sigma_j$ with probability 1 (simultaneously for each such pair i, j).

Many well known spin systems are monotone, including the Ising model and the hard-core model (independent sets).

2.5 Systems with hard constraints

Recall that according to our definition above, the edge potential U may only take on finite real values. However, there are interesting models where U is infinite for some combinations of spin values, i.e., there is a hard constraint forbidding certain combinations of spins along an edge. Examples of such systems are the hard-core model (whose configurations are independent sets) and the anti-ferromagnetic Potts model at zero temperature (whose configurations are proper colorings) - see e.g. [9] for definitions of these models. In general, the results of this paper apply to these kinds of systems as well. However, some of the notions we defined above are not necessarily well defined for systems with hard constraints. In order to avoid cumbersome details but still consider systems with hard constraints, we make the compromise of allowing U to be infinite but restricting our results to *permissive systems*. A permissive system is one where, for any finite subset Ψ and *any* boundary configuration τ, there is at least one configuration $\sigma \in \Omega_\Psi$ such that $H_\Psi^\tau(\sigma) < \infty$, and in particular, $\mu_\Psi^\tau(\sigma) > 0$. We also require that the space of "legal" configurations (those in the support of the stationary distribution) is connected under the Glauber dynamics. Notice that by definition, systems without hard constraints are always permissive. It is easy to verify that the hard-core model is permissive, as is the model of proper colorings when the number of colors is strictly larger than the degree of the lattice, i.e., $q > 2d$.

The main importance of assuming the system is permissive is that μ_Λ^τ is well defined for any value of τ. An alternative to this assumption is to extend the definition of μ_Λ^τ, but this requires additional details which we wish to avoid. Once the finite Gibbs distributions are well defined for any value of the boundary configuration, strong spatial mixing is also well defined. In addition, the transitions of the Markov chains above are well defined for any current configuration, even if it is not in the support of the stationary distribution. In permissive systems, the chain is guaranteed to reach a legal configuration at some finite time, and thus converge to the stationary Gibbs measure. Hence, without loss of generality, we may think of the chains as running on the whole configuration space Ω_Ψ. In particular, when we say the dynamics has optimal temporal mixing, the error bound is good for chains that start from illegal configurations as well. Notice, however, that this has a negligible quantitative effect since once every site is updated at least once (which takes $O(n \log n)$ time with high probability) the configuration is guaranteed to be a legal one.

2.6 Results

Several notions of temporal and spatial mixing for models on integer lattices are known to be equivalent to one another [4, 12, 13, 14, 17], though the proofs are often rather complex and cast in the language of functional analysis. In this paper we present combinatorial proofs of the following implications.

Theorem 2.3. *If the single-site dynamics has optimal temporal mixing then the system has strong spatial mixing.*

For monotone systems we show the converse as well:

Theorem 2.4. *If a monotone system has strong spatial mixing then the single-site dynamics has optimal temporal mixing.*

In the general case (without assuming monotonicity), we show that

Theorem 2.5. *If a system has strong spatial mixing then there exists a finite integer L for which the heat-bath block dynamics $HB(L)$ has optimal temporal mixing.*

The converse of Theorem 2.5 (that optimal temporal mixing of $HB(L)$ implies strong spatial mixing) can be proved using the same ideas as in the proof of Theorem 2.3 (with the addition of a few minor technical details), so we skip it here.

Notice that strong spatial mixing implies optimal temporal mixing of the single-site Glauber dynamics in the general case as well [4, 12, 14, 17], but we have not yet been able to find a purely combinatorial proof of this implication. The main obstacle is translating the rapid mixing result for the block dynamics into rapid mixing of the single-site dynamics (at the cost of only a constant factor), a problem which is still open for general spin-systems. The functional analysis proofs mentioned above analyze the log-Sobolev constant of the block dynamics, and the implication for the single-site dynamics follows since the log-Sobolev constant of the block dynamics translates easily to that of the single-site dynamics.

3 Preliminaries

In this section we identify some of the common tools we use in our proofs.

3.1 Coupling and Mixing Time

A common tool for bounding the total variation distance between two distributions, and in particular for bounding the mixing time of Markov chains, is coupling. A *coupling* of μ_1 and μ_2 is any joint distribution whose marginals are μ_1 and μ_2 respectively. If σ_1 and σ_2 are a pair of random configurations chosen from a given coupling of μ_1 and μ_2 then $\Pr(\sigma_1 \neq \sigma_2)$ is an upper bound on the total variation distance between μ_1 and μ_2. Also, there is always an optimal coupling, i.e., a coupling such that $\Pr(\sigma_1 \neq \sigma_2) = \|\mu_1 - \mu_2\|$.

In the proofs we give in this paper we use the following coupling of the Glauber dynamics, which we call an *identity coupling*. This coupling allows us to simultaneously couple any number of instances of the chain. An identity coupling is determined by specifying, for each site v, a coupling of all the single-site Gibbs distributions (ranging over all possible values for the configuration of the neighbors of v). Namely, we have a joint distribution γ_v whose marginals are $\mu^{\tau_1}_{\{v\}}, \ldots, \mu^{\tau_k}_{\{v\}}$, where the set $\{\tau_1, \ldots, \tau_k\} = \Omega_{\partial\{v\}}$. Given γ_v, we couple a collection of instances of the Glauber dynamics $(X^1_t), (X^2_t), \ldots, (X^l_t)$ using a Markovian coupling (i.e., the joint distribution of $X^1_{t+1}, \ldots, X^l_{t+1}$ is a function only of the coupled configurations $X^1_t, \ldots, X^l_t$) where the coupled transition $(X^1_t, \ldots, X^l_t) \to (X^1_{t+1}, \ldots, X^l_{t+1})$ is as follows:

- Choose a site v u.a.r. from Ψ (the same one for all chains).
- Choose a collection of spins $(s_1, \ldots, s_k)$ from the joint distribution γ_v.
- For every $1 \leq i \leq l$ set $X^i_{t+1}[v] = s_j$ if and only if $X^i_t[\partial\{v\}] = \tau_j$.

An important property of this coupling is that if $X^i_t[\partial\{v\}] = X^j_t[\partial\{v\}]$ then $X^i_{t+1}[v] = X^j_{t+1}[v]$ with probability 1. Notice that in a monotone system there exists a *monotone identity coupling*, i.e., a joint distribution γ_v such that whenever $\tau_i \succeq_{\partial\{v\}} \tau_j$, $s_i \succeq_v s_j$ with probability 1.

We say that an identity coupling has *optimal mixing* if for any two instances of the chain (X_t) and (Y_t), we have $\Pr(X_{kn} \neq Y_{kn}) \leq bn \exp(-ck)$, where the probability space is the coupling of X_{kn} and Y_{kn} resulting from the identity coupling of the two processes. Notice that optimal mixing of an identity coupling implies optimal temporal mixing of the dynamics. Finally, the *coupling time* of an identity coupling is the minimum T such that $\Pr(X_T \neq Y_T) \leq \frac{1}{e}$. As a result, $\Pr(X_{kT} \neq Y_{kT}) \leq e^{-k}$ for any positive integer k.

3.2 Bounding the Speed of Propagation of Information

A central idea in the analysis of the mixing time of the Glauber dynamics, in particular when using spatial mixing assumptions, is to bound the speed at which information propagates during the dynamical process. In this section we give a lemma of this sort following an argument explained to us by van den Berg, based on the idea of *paths of disagreement* (also known as *disagreement percolation* [2]). The idea of bounding the speed of propagation of information originally appeared in [17], and similar bounds can also be found in [12, 11]. The quantitative analysis in the argument in fact goes back to the Richardson model [16]. Our version below applies to the Glauber dynamics on general graphs of bounded degree (as in [11]), rather than just for finite subsets of $\mathbb{Z}^d$.

Lemma 3.1. *Let $G = (V, E)$ be a graph of maximum degree $\Delta > 1$, and let $n = |V|$. Let (X_t) and (Y_t) be two copies of a Glauber dynamics on G such that the two initial configurations agree everywhere except on $A \subseteq V$. Let $B \subseteq V$ be another subset and let $r = \mathrm{dist}(A, B)$. Then, for any positive integer $k \leq \frac{r}{(\Delta-1)e^2}$, if we run the dynamics for $T = kn$ steps, $\Pr(X_T[B] \neq Y_T[B]) \leq 4\min\{|A|, |B|\}\left(\frac{(\Delta-1)ek}{r}\right)^r$, where the probability space is the coupling of X_T and Y_T resulting from any identity coupling of (X_t) and (Y_t). In particular, if $T = kn$ and $\mathrm{dist}(A, B) \geq (\Delta-1)e^2k$, then $\Pr(X_T[B] \neq Y_T[B]) \leq 4\min\{|A|, |B|\}\, e^{-\mathrm{dist}(A,B)}$.*

In words, Lemma 3.1 states that in kn steps, with high probability, information percolates a distance of at most $(\Delta-1)e^2k$.

Proof: Since we couple X_t and Y_t using an identity coupling, if at time zero v had the same spin in both chains and at time T the spins at v differ then it must be the case that at some time $t' \leq T$ the site chosen to be updated was v and immediately before the update of v at time t' the two chains had different spins at one of the neighbors of v. Carrying this argument inductively, if we assume that at time zero the only sites whose spins may differ are included in A then in order for a site v to have different spins at time T there must be a *path of disagreement* going from A to v. Specifically, there must be $v_0, v_1, \ldots, v_l = v$ and $0 < t_1 < t_2 < \ldots < t_l \leq T$ such that $v_0 \in \Lambda$ and for $1 \leq i \leq l$, $v_i \sim v_{i-1}$ and at time t_i the site chosen to be updated was v_i. Notice that for a given path $v_0, \ldots, v_l$ the probability of this event occurring is at most $\binom{T}{l}(\frac{1}{n})^l$. Now, if the two configurations at time T differ at some site in B, there must be a path of disagreement of length at least $r = \mathrm{dist}(A, B)$ going from A to B. Since the number of (simple) paths of length l going from A to B is bounded from above by $\min\{|A|, |B|\}\,\Delta(\Delta-1)^{l-1}$ we can conclude that the probability of a disagreement in B at time $T = kn$ is at most

$$\begin{aligned}
\min\{|A|, |B|\} \cdot \frac{\Delta}{\Delta-1} \cdot \sum_{l=r}^{kn} (\Delta-1)^l \binom{kn}{l} \left(\frac{1}{n}\right)^l &\leq \min\{|A|, |B|\} \cdot \frac{\Delta}{\Delta-1} \cdot \sum_{l=r}^{\infty} \left(\frac{(\Delta-1)ek}{l}\right)^l \\
&\leq 4\min\{|A|, |B|\} \left(\frac{(\Delta-1)ek}{r}\right)^r,
\end{aligned}$$

where in the last inequality we used the fact that $r \geq (\Delta-1)e^2k$. ■

Remark: We will often use Lemma 3.1 in a setting where only a subset of the sites may be updated in the Markov chain (i.e., the spins on some sites - typically those on the boundary - are held fixed throughout the process). Notice that the proof above is still valid in this setting (regardless of whether or not the fixed spins disagree - i.e., are of sites in A). In fact, it is valid even if the two compared chains have different sets of fixed sites as long as the sites which are fixed in only one of the chains are all included in the subset A, i.e., we just assume that the spins of these sites disagree in the two chains. An important point to keep in mind in these scenarios is the meaning of the parameter n. Rather than the volume of the graph, n stands for the inverse of the probability that a given site is chosen to be updated (and it must be the same in both chains). Indeed, this is the only use we made of this parameter in the proof. The scenarios mentioned in this remark will become clearer when they arise in the proofs below.

4 From Temporal to Spatial Mixing

In this section we prove Theorem 2.3, which states that if the Glauber dynamics has optimal temporal mixing then strong spatial mixing holds. The first step in the proof is to derive a stronger notion of temporal mixing, given in Lemma 4.1 below. Temporal mixing as defined earlier (Definition 2.1) guarantees that if we run the dynamics on a rectangle Ψ for sufficient time then the distance between any two chains will be small enough as a function of the time we run the chains. The distance considered is the total variation distance between the two distributions on Ω_Ψ. However, if we *project* the distributions on Ω_Λ, where $\Lambda \subset \Psi$, it may very well be that after the same amount of time the distance between the two projected distributions is smaller than the distance between the original distributions. Ideally, we look for a bound which is of the same form as the one we get from running the dynamics on Λ, i.e., $b'|\Lambda|\exp(-c'k)$. We use the sub-exponential growth of $\mathbb{Z}^d$ to argue that if the Glauber dynamics has optimal temporal mixing then indeed this stronger notion, which we call *projected optimal mixing*, holds as well.

Lemma 4.1. *If the Glauber dynamics has optimal mixing then there exist constants b' and $c' > 0$ such that, for any subset Ψ of volume n, any boundary configuration, any two instances (X_t) and (Y_t) of the chain on Ψ and any subset $\Lambda \subseteq \Psi$, we have that $\|X_{kn} - Y_{kn}\|_\Lambda \leq b'|\Lambda|\exp(-c'k)$ for any positive integer k.*

Proof: The idea of the proof is one we use throughout this paper, which involves using Lemma 3.1 in order to localize the dynamics we consider. Namely, when we run the dynamics for kn steps, with high probability information from sites which are at distance at least $(2d-1)e^2k$ from Λ does not percolate into Λ. Therefore, if we take a subset Λ_k surrounding Λ and whose boundaries are at distance at least $(2d-1)e^2k$ from Λ, we can assume that the sites on the boundary of Λ_k are fixed throughout the process. Thus, we can use the optimal temporal mixing bound for a dynamics on the local subset Λ_k, whose volume is smaller than that of Ψ. As shown below, the fact that the volume of Λ_k grows only sub-exponentially in k (this is the first place where we use the sub-exponential growth of $\mathbb{Z}^d$) gives the required bound. An additional point we need to make in order to carry out the above argument is that when running the dynamics on Ψ, with high probability, an appropriate portion of the time is spent in the subset Λ_k. This, however, is an easy consequence of the Chernoff bound.

We proceed with the formal proof. Consider the subset of all sites within distance $(2d-1)e^2k$ from Λ, and let Λ_k be the intersection of this subset with Ψ. Notice that $\mathrm{dist}(\Lambda, \Psi\backslash\Lambda_k) \geq (2d-1)e^2k$ and that $|\Lambda_k| \leq [2(2d-1)e^2k]^d|\Lambda|$.

In addition to the chains (X_t) and (Y_t), we consider two additional chains, denoted by $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$, whose initial configurations inside Λ_k are the same as (X_t) and (Y_t) respectively. The configuration of $\Psi \setminus \Lambda_k$ is fixed to the same arbitrary configuration in both $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$ and remains fixed throughout the process, i.e., $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$ represent modified processes where, in a given step, if the chosen site to be updated is outside Λ_k then the spin of that site remains unchanged, while if it is in Λ_k then it is updated as usual. Notice that this modified process is the same as running the dynamics on Λ_k except that the probability of a site being chosen at a given step is $\frac{1}{|\Psi|}$ instead of $\frac{1}{|\Lambda_k|}$.

Using the triangle inequality, we have $\|X_{kn} - Y_{kn}\|_\Lambda \le \|X_{kn} - X_{kn}^{\Lambda_k}\|_\Lambda + \|X_{kn}^{\Lambda_k} - Y_{kn}^{\Lambda_k}\|_\Lambda + \|Y_{kn}^{\Lambda_k} - Y_{kn}\|_\Lambda$. Lemma 3.1 (together with the remark following it) gives a bound on the first and third terms in the r.h.s. of the last inequality. To see this, couple (X_t) and $(X_t^{\Lambda_k})$ using a modified identity coupling, where an update of a site outside Λ_k in (X_t) is coupled with doing nothing in $(X_t^{\Lambda_k})$. Notice that at time zero the two chains agree on Λ_k. Disagreement may percolate from $\Psi \setminus \Lambda_k$ into the bulk of Λ_k as we run the chains, but since $\mathrm{dist}(\Lambda, \Psi \setminus \Lambda_k) \ge (2d-1)e^2 k$, we can use Lemma 3.1 to deduce that $\|X_{kn} - X_{kn}^{\Lambda_k}\|_\Lambda \le 4|\Lambda| e^{-(2d-1)e^2 k}$.

It remains to bound $\|X_{kn}^{\Lambda_k} - Y_{kn}^{\Lambda_k}\|_\Lambda$. Recall that both these chains have the same fixed configuration outside Λ_k so we can use the optimal temporal mixing assumption for a process on Λ_k. Notice that when running the chain $X_t^{\Lambda_k}$ for kn steps, on average $k|\Lambda_k|$ of the steps hit Λ_k. Using a Chernoff bound, with probability at least $1 - \exp(-\frac{k|\Lambda_k|}{8})$, the number of steps that hit Λ_k is at least $\frac{k|\Lambda_k|}{2}$. Thus, we can use the same bound as when running a process on Λ_k for $\frac{k|\Lambda_k|}{2}$ steps. Specifically,

$$
\begin{aligned}
\|X_{kn}^{\Lambda_k} - Y_{kn}^{\Lambda_k}\|_\Lambda &\le \|X_{kn}^{\Lambda_k} - Y_{kn}^{\Lambda_k}\|_{\Lambda_k} \\
&\le b|\Lambda_k| \exp\left(-c \cdot \frac{k}{2}\right) + \exp\left(-\frac{k|\Lambda_k|}{8}\right) \\
&\le b[2(2d-1)e^2 k]^d |\Lambda| \exp\left(-c \cdot \frac{k}{2}\right) + \exp\left(-\frac{k}{8}\right) \\
&\le b'|\Lambda| \exp(-c' \cdot k)
\end{aligned}
$$

for appropriate constants b' and $c' > 0$. ■

We now proceed with the proof of Theorem 2.3.

Proof of Theorem 2.3: Let Ψ be a subset of volume n, τ and τ^u be two boundary configurations that differ only at u, and let $\Lambda \subseteq \Psi$. Following Lemma 4.1, we assume the dynamics has projected optimal mixing and show that

$$\|\mu_\Psi^\tau - \mu_\Psi^{\tau^u}\|_\Lambda \;\le\; b'|\Lambda| \exp\left(-\frac{c'}{(2d-1)e^2} \cdot \mathrm{dist}(u, \Lambda)\right) \;+\; 4|\Lambda| e^{-\mathrm{dist}(u,\Lambda)}.$$

The idea of the proof is that when running the Glauber dynamics, the time needed in order for the projected distribution on Λ to be close to the stationary one is less than the time it takes for the disagreement at u to percolate into Λ. Formally, consider the following two instances of the Glauber dynamics on Ψ. The first, denoted by Z_t, is an instance with τ as the boundary configuration while the second, denoted by Z'_t, is an instance with τ^u as the boundary configuration. The initial configuration of Ψ in both chains is chosen from the distribution $\mu_\Psi^{\tau^u}$. Notice that this is the stationary distribution of Z'_t and therefore $Z'_t = \mu_\Psi^{\tau^u}$ for all t.

Using the triangle inequality, we have $\|\mu_\Psi^\tau - \mu_\Psi^{\tau^u}\|_\Lambda = \|\mu_\Psi^\tau - Z_t'\|_\Lambda \le \|\mu_\Psi^\tau - Z_t\|_\Lambda + \|Z_t - Z_t'\|_\Lambda$. By letting $t = \frac{\mathrm{dist}(u,\Lambda)}{(2d-1)e^2} \cdot n$ we can make sure both terms are small. We bound the first term using the temporal mixing time assumption. Namely, for $t = \frac{\mathrm{dist}(u,\Lambda)}{(2d-1)e^2} \cdot n$ we have $\|\mu_\Psi^\tau - Z_t\|_\Lambda \le b'|\Lambda| \exp(-c' \cdot \frac{\mathrm{dist}(u,\Lambda)}{(2d-1)e^2})$. We use Lemma 3.1 in order to bound the second term. Notice that Z_t and Z_t' have the same initial distribution on Ψ and thus they can be coupled such that at time zero they have the same configuration on Ψ with probability 1. We continue to couple the two processes using an identity coupling. Disagreement may percolate from u, but since $\mathrm{dist}(u,\Lambda) = (2d-1)e^2 \frac{t}{n}$ we have $\|Z_t - Z_t'\|_\Lambda \le 4|\Lambda| e^{-\mathrm{dist}(u,\Lambda)}$. ■

We conclude this section with a couple of remarks on the generalization of the arguments made above to other settings. First, notice that we never used the fact that the difference on the boundary is only at a single site u. Indeed, if the difference is on a subset Δ we have the same bound (as a function of $\mathrm{dist}(\Delta,\Lambda)$) without adding any factor that depends on Δ. Second, the argument for showing that projected temporal mixing implies spatial mixing uses only Lemma 3.1 and can thus be carried out in models with any underlying finite degree graph. On the other hand, the proof of Lemma 4.1 uses the sub-exponential growth of $\mathbb{Z}^d$ and breaks down for graphs with exponential growth. Indeed, the Ising model on a tree at an appropriate temperature provides a counterexample to the claim of Lemma 4.1 in such graphs. This counterexample can be deduced from [11], where it is shown that there are temperatures where the Glauber dynamics for the Ising model on a tree has optimal temporal mixing but a modified form of strong spatial mixing (where the difference on the boundary may include many sites) does not hold, which in particular means that projected optimal mixing does not hold.

5 From Spatial to Temporal Mixing: The Monotone Case

In this section we show that in monotone systems the strong spatial mixing assumption implies optimal temporal mixing of the single-site Glauber dynamics (Theorem 2.4). Actually, we state two theorems whose combination gives Theorem 2.4. The first theorem, whose proof uses ideas from the proof of Theorem 4.2 of [13], states that the strong spatial mixing assumption implies $O(n \log^2 n)$ coupling time of any monotone identity coupling, uniformly in the volume n and in the boundary configuration. The second theorem, which is based on Theorem 3.12 of [12], states (for general systems) that if there exists n_0 for which the coupling time of any identity coupling of the Glauber dynamics on subsets of volume n_0 is at most $\frac{c}{\log n_0} n_0^{1+1/d}$ for an appropriate constant c, uniformly in the boundary configuration, then this identity coupling has optimal mixing. In particular, any upper bound of $o(\frac{n^{1+1/d}}{\log n})$ on the asymptotic coupling time immediately implies that the identity coupling has optimal mixing.

Theorem 5.1. *Strong spatial mixing implies that the coupling time of any monotone identity coupling of the Glauber dynamics on any subset of volume n is at most $T(n) = cn(\log n)^2$ for some constant c, uniformly in n and in the boundary configuration.*

Proof: As in our earlier arguments, the idea of the proof is again to localize the dynamics, which allows us to use inductive bounds from smaller volume subsets. However, here we use strong spatial mixing to achieve the localization, rather than the bound on the speed of propagation of information from Lemma 3.1.

Fix a large enough n_0 (to be determined later). By choosing an appropriate constant $c = c(n_0)$, the coupling time statement is true for all $n \leq n_0$. This is a consequence of the fact that any two instances of the chain will coalesce in finite time under any monotone coupling, e.g., because eventually both instances will simultaneously reach a maximal or minimal state. We go on to show the statement of the theorem is valid for $n > n_0$, by inductively assuming its validity for volumes $m \leq [2 \cdot \frac{2}{\alpha} \log(3e\beta n)]^d$, where α, β are the constants in the definition of strong spatial mixing (Definition 2.2).

Let Ψ be a subset of volume n with an arbitrary boundary configuration. Let (X_t) and (Y_t) be two instances of the chain with arbitrary initial configurations inside Ψ. We will show that after $T(n)$ steps, for every site $v \in \Psi$, the probability that the two spins at v differ is at most $\frac{1}{en}$, and therefore, the probability that two configurations (on the whole of Ψ) differ is at most $\frac{1}{e}$, as required.

Consider the regular box of radius $\frac{2}{\alpha} \log(3e\beta n)$ around v, and let Λ_v be the intersection of this box with Ψ. Let $m = |\Lambda_v|$ and notice that $m \leq [2 \cdot \frac{2}{\alpha} \log(3e\beta n)]^d$. We now introduce four additional chains that may only update sites in Λ_v. We will couple these chains along with (X_t) and (Y_t) such that, whenever the site chosen to be updated is outside Λ_v only X_t and Y_t are updated while the additional four chains remain unchanged. On the other hand, when the site to be updated belongs to Λ_v all six chains are updated simultaneously according to the monotone identity coupling. Below we describe the additional four chains and their initial configurations. Notice that we only describe the initial configuration inside Ψ. Outside Ψ, all four chains have the same boundary configuration as (X_t) and (Y_t).

1. Q_t^{+,Λ_v} : the chain starting from the all plus configuration on Ψ.

2. Q_t^{-,Λ_v} : the chain starting from the all minus configuration on Ψ.

3. Z_t^{+,Λ_v} : the chain starting from the all plus configuration outside Λ_v, while the initial configuration inside Λ_v is chosen from the (stationary) Gibbs measure on Λ_v with this boundary configuration.

4. Z_t^{-,Λ_v} : the chain starting from the all minus configuration outside Λ_v, and the stationary Gibbs measure corresponding to this boundary configuration inside Λ_v.

Notice that we can simultaneously couple the six chains such that at time zero, with probability one, $Q_0^{+,\Lambda_v} \succeq X_0 \succeq Q_0^{-,\Lambda_v}$, $Q_0^{+,\Lambda_v} \succeq Y_0 \succeq Q_0^{-,\Lambda_v}$, and $Z_t^{+,\Lambda_v} \succeq Z_t^{-,\Lambda_v}$. Since we use a monotone identity coupling, we have by induction that these relations hold for all t. Thus, we have

$$\Pr(X_t[v] \neq Y_t[v]) \;\leq\; \Pr(Q_t^{+,\Lambda_v}[v] \neq Q_t^{-,\Lambda_v}[v]) \;\leq$$
$$\Pr(Q_t^{+,\Lambda_v}[v] \neq Z_t^{+,\Lambda_v}[v]) \;+\; \Pr(Z_t^{+,\Lambda_v}[v] \neq Z_t^{-,\Lambda_v}[v]) \;+\; \Pr(Z_t^{-,\Lambda_v}[v] \neq Q_t^{-,\Lambda_v}[v]).$$

We use the strong spatial mixing assumption to bound the middle term of the last expression. Notice that since Z_t^{+,Λ_v} and Z_t^{-,Λ_v} represent the stationary Gibbs distributions on Λ_v with the appropriate boundary configurations then strong spatial mixing (together with the triangle inequality[4]) gives $\|Z_t^{+,\Lambda_v} - Z_t^{-,\Lambda_v}\|_{\{v\}} \leq |\partial\Lambda_v \setminus \partial\Psi| \beta \exp(-\alpha \cdot \mathrm{dist}(\partial\Lambda_v \setminus \partial\Psi, v))$. This bound on the

[4]The strong spatial mixing assumption gives bounds only for comparing two Gibbs distributions whose boundary conditions differ at a single site. We use the triangle inequality in order to extend the bound to comparing two distributions whose boundary conditions differ at multiple sites.

total variation distance does not guarantee the same bound on disagreement under the coupling because the coupling we use is not necessarily the optimal one. However, monotonicity guarantees that our coupling is within a factor of $q-1$ (recall that q is the size of the spin space) from the optimal coupling, as explained next. We embed the ordering associated with v in the linear ordering $1, 2, \ldots, q$ with integer arithmetic. Since the spins at v are coupled such that with probability one $Z_t^{+,\Lambda_v}[v] \geq Z_t^{-,\Lambda_v}[v]$, we have

$$
\begin{aligned}
\Pr(Z_t^{+,\Lambda_v}[v] \neq Z_t^{-,\Lambda_v}[v]) &\leq \mathrm{E}(Z_t^{+,\Lambda_v}[v] - Z_t^{-,\Lambda_v}[v]) \\
&= \mathrm{E}(Z_t^{+,\Lambda_v}[v]) - \mathrm{E}(Z_t^{-,\Lambda_v}[v]) \\
&\leq (q-1)\|Z_t^{+,\Lambda_v} - Z_t^{-,\Lambda_v}\|_{\{v\}} \\
&\leq (q-1)|\partial\Lambda_v \setminus \partial\Psi|\beta\exp(-\alpha \cdot \mathrm{dist}(\partial\Lambda_v \setminus \partial\Psi, v)) \\
&\leq \frac{1}{3en}
\end{aligned}
$$

for large enough n. Notice that in order to get the inequality in the middle line we used an optimal coupling of $Z_t^{+,\Lambda_v}[v]$ and $Z_t^{-,\Lambda_v}[v]$ together with the fact that the oscillation of any function whose range is $[1, q]$ is at most $q-1$.

In order to complete the proof we have to show that $\Pr(Q_t^{+,\Lambda_v}[v] \neq Z_t^{+,\Lambda_v}[v]) \leq \frac{1}{3en}$ when $t = T(n)$ (by symmetry, the same will hold for the minus chains). Using a Chernoff bound, if we run the dynamics on Ψ for $cn(\log n)^2$ steps then with probability at least $1 - \frac{1}{6en}$ the number of steps in which Λ_v is hit is at least

$$
\frac{1}{2}cm(\log n)^2 = (2\log n)cm\frac{\log n}{4} \geq (2\log n)cm(\log m)^2
$$

for large enough n. If we assume that indeed Λ_v is hit this often then we can use the induction hypothesis to bound the probability that the spins at v differ because the two chains we are comparing have the same fixed boundaries outside Λ_v. Indeed, after $T(m) = cm(\log m)^2$ steps in Λ_v, the configurations (on the whole of Λ_v) disagree with probability at most $\frac{1}{e}$, and thus after $(2\log n)T(m)$ steps, they disagree with probability at most $\frac{1}{n^2}$. Hence, $\Pr(Q_{T(n)}^{+,\Lambda_v}[v] \neq Z_{T(n)}^{+,\Lambda_v}[v]) \leq \frac{1}{6en} + \frac{1}{n^2} \leq \frac{1}{3en}$ for large enough n, as required. ■

Remark: The reader may have noticed that, by carrying through a more careful analysis in the above proof, one can get a slightly better bound — for example, $O(n\log n(\log\log n)^2)$ — on the coupling time. However, since in any case we will reduce the coupling time to $O(n\log n)$ using the next theorem, we choose to keep the calculations simpler by only showing a bound of $O(n\log^2 n)$.

Theorem 5.2. *Suppose there exists an identity coupling such that for all subsets Λ of volume at most n_0, where n_0 is a sufficiently large constant, the coupling time of the given identity coupling on Λ is at most $\frac{1}{8(2d-1)e^2}\frac{n_0^{1/d}}{\log n_0}|\Lambda|$ uniformly in the boundary configuration. Then for all n and for all subsets Ψ of volume n with any boundary configuration, $\Pr(X_{kn} \neq Y_{kn}) \leq |\Psi|\exp(-ck)$, where $c = 2(2d-1)e^2 n_0^{-\frac{1}{d}}$. Namely, this identity coupling has optimal mixing.*

Proof: Consider the Glauber dynamics on Ψ with an arbitrary boundary configuration. We will show that for any two instances of the chain (X_t) and (Y_t) and any $v \in \Psi$ we have $\Pr(X_{kn}[v] \neq$

$Y_{kn}[v]) \le \exp(-ck)$ under the given identity coupling. Using a union bound, this implies that $\Pr(X_{kn} \ne Y_{kn}) \le |\Psi| \exp(-ck)$.

Let $l_0 = \lceil \frac{1}{c} \rceil = \lceil \frac{n_0^{1/d}}{2(2d-1)e^2} \rceil$. As before, we will use Lemma 3.1 to localize the dynamics. Together with the hypothesis of the theorem, this will imply that after $l_0 n$ steps the spins at v agree with high probability. What we want, however, is that the probability of disagreement will continue to decay exponentially with the number of steps. Notice that such a result would follow if, once the spins at v agreed, they continued to agree through the rest of the process, but this is clearly not the case. However, using the sub-exponential growth of $\mathbb{Z}^d$ and another localization argument, we can show that if all the spins within a large enough radius around v agree at a given time, then the spins at v will continue to agree for sufficiently many steps (depending on the radius of agreement). Bootstrapping from the sufficiently small probability of disagreement after $l_0 n$ steps, we get the required exponential decay.

We proceed with the formal proof. Let $\rho(k) = \max_{X_0, Y_0, v \in \Psi} \Pr(X_{kn}[v] \ne Y_{kn}[v])$. We have the following two claims.

Claim 1. *Under the hypothesis of the theorem,* $\rho(l_0) \le \frac{1}{e2^d(n_0+1)} = \frac{1}{e2^d([2(2d-1)e^2 l_0]^d+1)}$.

Claim 2. *Without any assumptions, for any* k_1 *and* k_2, $\rho(k_1+k_2) \le [2(2d-1)e^2 k_2]^d \rho(k_1)\rho(k_2) + 4e^{-k_2}$.

Theorem 5.2 follows from the combination of the above two claims. To see this, let $\phi(k) = 2^d([2(2d-1)e^2 k]^d + 1) \cdot \max\left\{\rho(k), 2e^{-\frac{k}{2}}\right\}$. Using Claim 2, we have by an explicit calculation that $\phi(2k) \le \phi(k)^2$. On the other hand, from Claim 1 we get that $\phi(l_0) \le \frac{1}{e}$ (where we used the fact that l_0 is large enough to handle the case of $\rho(l_0) < 2e^{-\frac{l_0}{2}}$). We then conclude that $\rho(k) \le \phi(k) \le \exp(-\frac{k}{l_0})$, as required.

Proof of Claim 1: Let $v \in \Psi$ be any site. As in Lemma 4.1, the idea is to take a regular box of volume n_0 around v. Then, since we run the coupled chains for only $l_0 n$ steps, information from the boundary of this box does not have enough time to percolate to v. We can therefore assume the boundaries around this box are fixed. But then, the assumption of the theorem guarantees that the spins at v will agree with the required probability.

Formally, let Λ_v be the intersection of the regular box of volume n_0 around v with Ψ. Let $(X_t^{\Lambda_v})$ and $(Y_t^{\Lambda_v})$ be two chains whose initial configurations inside Λ_v agree with X_0 and Y_0 respectively, and which have the same fixed arbitrary boundary configuration on $\partial\Lambda_v \setminus \partial\Psi$. We have $\Pr(X_t[v] \ne Y_t[v]) \le \Pr(X_t[v] \ne X_t^{\Lambda_v}[v]) + \Pr(X_t^{\Lambda_v}[v] \ne Y_t^{\Lambda_v}[v]) + \Pr(Y_t^{\Lambda_v}[v] \ne Y_t[v])$. Notice that $\mathrm{dist}(v, \partial\Lambda_v \setminus \partial\Psi) \ge \frac{1}{2} n_0^{1/d} = (2d-1)e^2 l_0$. Therefore, using Lemma 3.1, we have $\Pr(X_{l_0 n}[v] \ne X_{l_0 n}^{\Lambda_v}[v]) \le 4e^{-(2d-1)e^2 l_0}$.

We go on to bound $\Pr(X_{l_0 n}^{\Lambda_v}[v] \ne Y_{l_0 n}^{\Lambda_v}[v])$. Notice that since in both chains the configuration outside Λ_v is fixed and is identical in both chains and since $|\Lambda_v| \le n_0$, we can use the hypothesis of the theorem to bound the above probability. If we run the coupled chains for $l_0 n$ steps, then with probability at least $1 - \exp(-\frac{l_0}{8}|\Lambda_v|)$ the number of steps that hit Λ_v is at least $\frac{l_0}{2}|\Lambda_v|$. If indeed that many steps hit Λ_v then according to the hypothesis of the theorem, $\Pr(X_t^{\Lambda_v}[v] \ne Y_t^{\Lambda_v}[v]) \le e^{-2\log n_0} = n_0^{-2}$. Thus, $\Pr(X_{l_0 n}^{\Lambda_v}[v] \ne Y_{l_0 n}^{\Lambda_v}[v]) \le n_0^{-2} + \exp(-\frac{l_0}{8}|\Lambda_v|)$. Putting this together with the result of the previous paragraph we get $\Pr(X_{l_0 n}[v] \ne Y_{l_0 n}[v]) \le n_0^{-2} + \exp(-\frac{l_0}{8}) + 8e^{-(2d-1)e^2 l_0} \le \frac{1}{e2^d(n_0+1)}$ for sufficiently large n_0, as required. $\square$

Proof of Claim 2: We use Lemma 3.1 once again, this time in the sense that in k_2n steps, information can percolate over a distance of at most $(2d-1)e^2k_2$. Thus, if the spins of all the sites within that radius from v agree after k_1n steps, then the spin at v will continue to agree after $(k_1+k_2)n$ steps with high probability.

Formally, let Λ_{v,k_2} be the intersection of the regular box of radius $(2d-1)e^2k_2$ around v with Ψ, and let $\mathcal{A}$ stand for the event that $X_{k_1n}[\Lambda_{v,k_2}] \neq Y_{k_1n}[\Lambda_{v,k_2}]$. Then, using Lemma 3.1 we have

$$\Pr(X_{(k_1+k_2)n}[v] \neq Y_{(k_1+k_2)n}[v]) \leq (1-\Pr(\mathcal{A}))4e^{-(2d-1)e^2k_2} + \Pr(\mathcal{A})\rho(k_2).$$

The proof is concluded once we notice that $\Pr(\mathcal{A}) \leq |\Lambda_{v,k_2}|\rho(k_1) \leq [2(2d-1)e^2k_2]^d\rho(k_1)$. □

This completes the proof of Theorem 5.2. ■

Remark: Notice that, in fact, the proof of Theorem 5.2 gives the stronger property of projected optimal mixing, as in Lemma 4.1. The hypothesis of Theorem 5.2 differs from that of Lemma 4.1 in two respects. On one hand, the hypothesis of Theorem 5.2 is stronger because it works with the coupling time of an identity coupling rather than with the mixing time in general. On the other hand, the hypothesis in Theorem 5.2 is weaker because the time bounds are weaker. The reason why a weaker time bound is sufficient for coupling time is that we can appeal to the union bound $\Pr(X_t[\Lambda] \neq Y_t[\Lambda]) \leq \sum_{v\in\Lambda} \Pr(X_t[v] \neq Y_t[v])$. We used this union bound twice, first when we reduced the proof to bounding the probability of disagreement at a single site, and second when we bounded the probability of the event $\mathcal{A}$. Notice that the corresponding inequality for the total variation distance is not necessarily true. Namely, we cannot in general assert that $\|X_t - Y_t\|_\Lambda \leq \sum_{v\in\Lambda} \|X_t - Y_t\|_{\{v\}}$. If this assertion were true then we could have done with assuming a fast mixing time (rather than a fast coupling time) and working with the total variation distance rather than with the probability of disagreement throughout the proof.

As remarked at the beginning of this section, combining Theorems 5.1 and 5.2 immediately yields Theorem 2.4.

6 From Spatial to Temporal Mixing: The General Case

In this section we prove Theorem 2.5. Namely, we show that in general (without assuming monotonicity), strong spatial mixing implies that the heat-bath block dynamics has optimal temporal mixing if the blocks used are large enough. Using *path coupling* [3], the proof is reduced to showing that strong spatial mixing implies that the so-called *Dobrushin-Shlosman condition* for complete analyticity [6] holds. This last implication was proven in [6], but we include a simple proof of it here.

Proof of Theorem 2.5: Consider the heat-bath dynamics $HB(L)$ on a rectangle Ψ of volume n with an arbitrary boundary configuration. Notice that L here is a large enough constant to be set later and will depend only on the dimension d and the constants from the definition of strong spatial mixing. In particular, L is uniform in n and the boundary configuration. Recall that the dynamics chooses a block to be updated from $B(\Psi, L)$, which is the set of translations of the regular box of side-length L that intersect Ψ. We denote the number of blocks by $m = |B(\Psi, L)|$ and notice that

$n \le m \le L^d n$ (the lower bound is due to the fact that the number of translations that intersect Ψ is at least the volume of Ψ while the upper bound crudely uses the fact that each site is covered by L^d translations). Using the *path coupling* method [3], it is enough to show that there exists a constant $c > 0$ (independent of n and the boundary configuration) such that, for any site $u \in \Psi$ and any two configurations σ,σ^u that differ only at u, there exists a coupling of the two chains whose current configurations are σ and σ^u respectively such that after one step, the average *Hamming* distance between the two coupled configurations is at most $1 - \frac{c}{m}$, i.e, decreases by at least $\frac{c}{m}$.

We couple these two chains using a specific identity coupling. Namely, the block chosen to be updated is the same in both chains, and if the boundaries of that block are the same in both σ and σ^u then we couple the update of the block such that the configurations inside the block agree with probability one. If the boundaries are not the same (this can happen only if u is on the boundary of the chosen block), we use a coupling to be described below.

From the way we defined the heat-bath block dynamics, each site in Ψ is included in exactly L^d blocks. Since we use an identity coupling, if a block including the site u is chosen to be updated then the Hamming distance between the two configurations will be zero (i.e., decrease by one) since the boundaries of this block are the same in both σ and σ^u. The probability of choosing a block as above is $\frac{L^d}{m}$. Thus, it is enough to show that the contribution to the expected change in Hamming distance from choosing the rest of the blocks is at most $\frac{L^d - c}{m}$.

As we already mentioned, the Hamming distance may increase only if the block chosen to be updated is one whose boundaries include u. Since there are at most $2dL^{d-1}$ such blocks, we will be done once we show that we can couple the update of each such block Λ such that the resulting average Hamming distance in Λ is strictly less then $\frac{L}{2d}$.

Consider a block Λ such that $u \in \partial\Lambda$. Let $r = \frac{1}{2}(\frac{L}{4d})^{\frac{1}{d}}$, $\Lambda_r = \{v \in \Lambda \,|\, \mathrm{dist}(v,u) \le r\}$, and $\overline{\Lambda_r} = \Lambda \setminus \Lambda_r$. By the strong spatial mixing assumption, $\|\mu_\Lambda^\sigma - \mu_\Lambda^{\sigma^u}\|_{\overline{\Lambda_r}} \le \beta|\overline{\Lambda_r}|\exp(-\alpha \cdot r) \le L^{-d}$ for a large enough L. We can thus couple the update of Λ such that the two coupled configurations disagree over $\overline{\Lambda_r}$ with probability at most L^{-d}. A trivial upper bound on the resulting average Hamming distance in Λ in this coupling is then $|\Lambda_r| + L^{-d}|\overline{\Lambda_r}| \le \frac{L}{4d} + 1$. ■

References

[1] M. Aizenman and R. Holley, Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime, in: H. Kesten, *Percolation Theory and Ergodic Theory of Infinite Particle Systems*, IMS Volumes in Math. and Appl. **8**, 1-11, Springer-Verlag, New York (1987).

[2] J. van den Berg, A uniqueness condition for Gibbs measures, with applications to the 2-dimensional Ising antiferromagnet, *Comm. Math. Phys.*, **152**, 161-166 (1993).

[3] R. Bubley and M.E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, 223-231 (1997).

[4] F. Cesi, Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields, *Probab. Theory and Related Fields*, **120**, 569-584 (2001).

[5] F. Cesi, G. Guadagni, F. Martinelli and R.H. Schonmann, On the 2D stochastic Ising model in the phase coexistence region close to the critical point, *Journ. Stat. Phys.* **85**, 55-102 (1996).

[6] R.L. Dobrushin and S.B. Shlosman, Completely Analytical Gibbs Fields, in: J. Fritz, A. Jaffe, D. Szasz, *Statistical Mechanics and Dynamical Systems*, 371-403, Birkhauser, Boston (1985).

[7] J.A. Fill and M. Machida, Stochastic monotonicity and realizable monotonicity, *Annals of Probability*, **29**, 938-978 (2001).

[8] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter, Berlin (1988).

[9] H.-O. Georgii, O. Häggström and C. Maes, The random geometry of equilibrium phases, *Phase Transitions and Critical Phenomena*, Volume 18 (C. Domb and J.L. Lebowitz, eds), 1-142, Academic Press, London (2001).

[10] R. Holley, Possible rates of convergence in finite range, attractive spin systems, in: R. Durret, *Particle systems, Random media, and Large deviations*, AMS Volumes on Contemp. Math. **41**, 215-234, (1985).

[11] C. Kenyon, E. Mossel and Y. Peres, Glauber dynamics on trees and hyperbolic graphs, *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, 568-578 (2001).

[12] F. Martinelli, Lectures on Glauber dynamics for discrete spin models, *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, Lecture notes in Math. **1717**, 93-191, Springer, Berlin (1998).

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

[14] F. Martinelli and E. Olivieri, Approach to equilibrium of Glauber dynamics in the one phase region II: The general case, *Comm. Math. Phys.* **161**, 487-514 (1994).

[15] F. Martinelli, E. Olivieri and R.H. Schonmann, For Gibbs state of 2D lattice spin systems weak mixing implies strong mixing, *Comm. Math. Phys.* **165**, 33-47 (1994).

[16] D. Richardson, Random growth in a tesselation, *Proc. Cambr. Phil. Soc.* **74**, 515-528 (1973).

[17] D.W. Stroock and B. Zegarlinski, The logarithmic Sobolev inequality for discrete spin systems on a lattice, *Comm. Math. Phys.* **149**, 175-194 (1992).

[18] B. Zegarlinski, On log-Sobolev inequalities for infinite lattice systems, *Lett. Math. Phys.* **20**, 173-182 (1990).