Commun. Math. Phys. 323, 1007–1016 (2013)
Digital Object Identifier (DOI) 10.1007/s00220-013-1776-0

Communications in
**Mathematical
Physics**

Can Extra Updates Delay Mixing?

Yuval Peres[1], **Peter Winkler**[2]

[1] Microsoft Research, Redmond, WA 98052, USA. E-mail: peres@microsoft.com
[2] Dept. of Mathematics, Dartmouth College, Hanover, NH 03755, USA

Received: 6 May 2012 / Accepted: 12 February 2013
Published online: 31 August 2013 –

Abstract: We consider Glauber dynamics (starting from an extremal configuration) in a monotone spin system, and show that interjecting extra updates cannot increase the expected Hamming distance or the total variation distance to the stationary distribution. We deduce that for monotone Markov random fields, when block dynamics contracts a Hamming metric, single-site dynamics mixes in $O(n \log n)$ steps on an n-vertex graph. In particular, our result completes work of Kenyon, Mossel and Peres concerning Glauber dynamics for the Ising model on trees. Our approach also shows that on bipartite graphs, alternating updates systematically between odd and even vertices cannot improve the mixing time by more than a factor of $\log n$ compared to updates at uniform random locations on an n-vertex graph. Our result is especially effective in comparing block and single-site dynamics; it has already been used in works of Martinelli, Toninelli, Sinclair, Mossel, Sly, Ding, Lubetzky, and Peres in various combinations.

1. Introduction

In a number of cases, mixing rates have been determined for Glauber dynamics using block updates, but only rough estimates have been obtained for single site dynamics. Examples include the Ising model on trees and the monomer-dimer model on $\mathbb{Z}^d$. In this work, we employ a "censoring lemma" for monotone systems in order to transport bounds for block dynamics to bounds for single site dynamics; sharp estimates result in several situations.

Our main interest is in ferromagnetic spin systems with nearest-neighbor interactions on a finite graph G. A configuration consists of a mapping σ from the set V of sites of G to a fixed partially ordered set S of "spins". The probability $\pi(\sigma)$ of a configuration σ is given by

$$\frac{1}{Z} \prod_{u \sim v} \Psi(\sigma_u, \sigma_v),$$

where Z is the appropriate normalizing constant. More generally, our results apply when π defines a monotone Markov random field. In single site Glauber dynamics, at each step, a uniformly random site is "updated" and assumes a new spin according to π conditioned on the spins of its neighbors. The resulting Markov chain is irreducible, aperiodic, and has unique stationary distribution π. Let $p^t(\omega, \cdot)$ be the distribution of configurations after t steps, with initial state ω. Let $\|\mu - \nu\| = \frac{1}{2}\sum_\sigma |\mu(\sigma) - \nu(\sigma)|$ be the total variation distance. The *mixing time* $T_G(\epsilon, \omega)$ for initial state ω is the least t such that $\|p^t(\omega, \cdot) - \pi\| \le \epsilon$. Finally, the (overall) **mixing time** $T_G(\epsilon)$ for the dynamics is $\max_{\omega\in\Omega} T_G(\epsilon, \omega)$.

In discrete-time block dynamics, a family $\mathcal{B}$ of "blocks" of sites is provided. At each step, a block $B \in \mathcal{B}$ is selected uniformly at random and a configuration on B is selected according to π conditioned on the spins of the sites in the exterior boundary of B. A useful method of bounding mixing times is to first bound the spectral gap of the block dynamics using path coupling, and then use comparison theorems for the spectral gap to derive a bound for $T_G(\epsilon)$. In key examples of Glauber dynamics for the Ising model on lattices and trees, this method tends to overestimate $T_G(\epsilon)$ by a factor of n on an n-vertex graph.

Stated informally, our main results are:

- In Glauber dynamics for a monotone (i.e., attractive) spin system, started at the top or bottom state, censoring updates increases the distance from stationarity.
- Suppose a monotone spin system on an n-vertex graph G has a block dynamics which contracts (on average) a weighted Hamming metric (see the remark following Theorem 1.1), and single-site dynamics on each block with arbitrary boundary conditions mixes in a bounded time. If the collection of blocks can be partitioned into a bounded number of layers such that blocks in each layer are nonadjacent, and weights within a block have a bounded ratio, then discrete time single site dynamics on G mixes (in total variation) in $O(n \log n)$ steps.
- In [12] (see also [3]) it was proved that for the Ising model on an n-vertex b-ary tree, block dynamics with large bounded blocks contracts a (weighted) Hamming metric at temperatures above the extremality threshold. This, in conjunction with our main results, implies that single-site dynamics on these trees mixes in $O(n \log n)$ steps. (See [19] for refinements of this theorem using Log-Sobolev inequalities).
- If H is a subgraph of G and only one vertex in H is adjacent to vertices in $G\backslash H$, then continuous-time Glauber dynamics on H mixes faster than the restriction to H of continuous-time Glauber dynamics on G.
- For an n vertex bipartite graph, alternating updating of all the "odd" and all the "even" vertices cannot mix much faster than systematic updates (enumerating the vertices in an arbitrary order): The odd-even updates can reduce the number of vertices updated at most by a factor of two. Similarly, the odd-even updates can be faster than uniformly random updates by a factor of at most $\log n$.

See §1.2 for further discussion of block dynamics, and §2-3 for proofs. A preliminary version of our results, including the proof of Theorem 1.1, was presented in the 2005 lectures [23].

1.1. Terminology. In what follows, a *system* $\langle \Omega, S, V, \pi \rangle$ consists of a finite set S of spins, a set V of sites, a space $\Omega \subseteq S^V$ of configurations (assignments of spins to sites), and a distribution π on Ω, which will serve as the stationary distribution for our Glauber dynamics. We assume that $\pi(\omega) > 0$ for $\omega \in \Omega$. The Ising model (where $S = \{+, -\}$

and $\Omega = S^V$) is the basic example; we allow Ω to be a strict subset of S^V to account for "hard constraints" such as those imposed by the hard-core gas model.

We denote by σ_v^s the configuration obtained from σ by changing its value at v to s, that is, $\sigma_v^s(v) = s$ and $\sigma_v^s(u) = \sigma(u)$ for all $u \neq v$. Let $\sigma_v^\bullet$ be the set of configurations $\{\sigma_v^s\}_{s \in S}$ in Ω.

The **update** μ_v at v of a distribution μ on Ω is defined by

$$\mu_v(\sigma) = \frac{\pi(\sigma)}{\pi(\sigma_v^\bullet)}\mu(\sigma_v^\bullet) \quad \text{for } \sigma \in \Omega. \tag{1}$$

For measures μ and ν on a poset Γ, we write $\nu \preceq \mu$ to indicate that μ *stochastically dominates* ν, that is, $\int g\, d\nu \leq \int g\, d\mu$ for all increasing functions $g : \Gamma \to \mathbb{R}$.

The system $\langle \Omega, S, V, \pi \rangle$ is called **monotone** if S is totally ordered, S^V is endowed with the coordinate-wise partial order, and whenever $\sigma, \tau \in \Omega$ satisfy $\sigma \leq \tau$, then for any vertex $v \in V$ we have

$$\left\{ \frac{\pi(\sigma_v^s)}{\pi(\sigma_v^\bullet)} \right\}_{s \in S} \preceq \left\{ \frac{\pi(\tau_v^s)}{\pi(\tau_v^\bullet)} \right\}_{s \in S} \tag{2}$$

as distributions on the spin set S. Let + denote the maximal element of S; we will assume that the all "+" configuration is in Ω, and refer to it as *the top configuration.*

1.2. Main results.

Theorem 1.1. *Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system and let μ be the distribution on Ω which results from successive updates at sites $v_1, \dots, v_m$, beginning at the top configuration. Let ν be defined similarly but with updates only at a subsequence $v_{i_1}, \dots, v_{i_k}$. Then $\mu \preceq \nu$, and $\|\mu - \pi\| \leq \|\nu - \pi\|$ in total variation. Moreover, this also holds if the sequence $v_1, \dots, v_m$ and its subsequence $i_1, \dots, i_k$ are chosen at random according to any prescribed distribution.*

See §2 for the proof, which shows also that the assumption of starting from the top configuration can be replaced by the assumption that the dynamics starts at a distribution μ_0, where the likelihood ratio μ_0/π is weakly increasing. Other assumptions, in particular monotonicity of the system, cannot be dispensed with, as shown recently by Holroyd [11].

Remark. Fix positive weights $\{w_v\}_{v \in V}$. Note that in the binary case $S = \{0, 1\}$, the average weighted Hamming distance in a monotone coupling between μ and π (when $\pi \preceq \mu$) can be written as $\Upsilon_w(\mu) - \Upsilon_w(\pi)$, where $\Upsilon_w(\mu) = \sum_{\sigma \in \Omega} \mu(\sigma) \sum_{v \in V} w_v \sigma(v)$. Since $\mu \preceq \nu$ implies that $\Upsilon_w(\mu) \leq \Upsilon_w(\nu)$, this justifies the assertion about Hamming distance in the abstract.

Next, we discuss block dynamics and the contraction method to bound mixing times for spin systems.

Let us endow $\Omega \subset S^V$ with the Hamming metric $H(\sigma, \tau) = |\{v \in V : \sigma_v \neq \tau_v\}|$. (More generally, it is sometimes fruitful to consider a weighted ℓ^1 metric). The **Kantorovich (transportation) distance** $\rho(\mu, \nu)$ between two distributions on Ω is defined to be the minimum over all couplings of μ and ν of $\mathbf{E}H(\sigma, \tau)$, where σ is drawn from μ and τ from ν. The fact that this metric satisfies the triangle inequality is proved, e.g., in Chap. 14 of [13] and is essentially equivalent to the path-coupling theorem of [4].

Given a subset B of V, let $\sigma_B^\bullet$ be the set of configurations $\tau \in \Omega$ such that τ agrees with σ on $V\backslash B$. For $\sigma \in \Omega$, the *block update* $U_B\sigma$ is a measure on $\sigma_B^\bullet$ defined by $(U_B\sigma)(\omega) = \frac{\pi(\omega)}{\pi(\sigma_B^\bullet)}$ for $\omega \in \sigma_B^\bullet$. Thus $U_B\sigma$ is π conditioned on $\sigma_B^\bullet$. For a collection of blocks $\mathcal{B}$, the $\mathcal{B}$-*averaged block update* of $\sigma \in \Omega$ yields a random configuration with distribution $\frac{1}{|\mathcal{B}|}\sum_{B\in\mathcal{B}} U_B\sigma$. The *block dynamics* determined by $\mathcal{B}$ consists of performing successive $\mathcal{B}$-averaged block updates. For simplicity, we will assume that all blocks in $\mathcal{B}$ have the same size.

We say that a block dynamics is *contracting* if for any two configurations σ and τ, the average Kantorovich distance after an update of a random block satisfies

$$\frac{1}{|\mathcal{B}|}\sum_{B\in\mathcal{B}} \rho(U_B\sigma, U_B\tau) \le \left(1-\gamma\frac{|B|}{|V|}\right) H(\sigma,\tau), \tag{3}$$

where γ is a constant and $|B|$ is the number of sites in a block. The triangle inequality for the Kantorovich metric implies that it suffices to verify this contraction condition when σ and τ differ at a single site. In our setting, contraction implies a bound of order $|V|\log|V|$ on the mixing time of the block dynamics, since the number of blocks is of order $|V|$. When the blocks are cubes in a lattice, a sufficient condition for contraction of block dynamics is *strong spatial mixing*, as defined and studied in [9,15–17].

Next, suppose that V is the vertex set of a graph. The system $\langle\Omega, S, V, \pi\rangle$ is a **Markov random field** if for any set $B \subset V$ and $\sigma \in \Omega$, the distribution $U_B\sigma$ depends only on the restriction of σ to ∂B, the set of vertices in B^c that are adjacent to B.

The next theorem is intended to illustrate how, in a particular case, Theorem 1.1 can be used to deduce rapid mixing for single-site dynamics from a contraction condition for block dynamics.

Theorem 1.2. *Let Ω be the configuration space for a monotone Markov random field on the d-dimensional toroidal grid $V = [0, N-1]^d$. Suppose that $\ell > d$ is odd and $\ell + 1$ divides N. Let $\mathcal{B} = \{B_v : v \in V\}$, where B_v is the cube of side-length $\ell - 1$ centered at v (so the cardinality of B is ℓ^d). If the block dynamics determined by $\mathcal{B}$ is contracting, and the single-site dynamics restricted to any block B has mixing time $T_B(1/4) \le t_0$ (for all boundary conditions), then single-site dynamics on all of V has mixing time* $\mathrm{O}(|V|\log|V|)$*, where the implied constant depends only on the contraction parameter γ and on ℓ, t_0 and d.*

Proof. For any $u \in V$ and any $s \in S$, suppose $\sigma' = \sigma_u^s$ is obtained from σ by changing the spin at u to s. Since $H(\sigma, \sigma') = 1$, we have $\rho(U_B\sigma, U_B\sigma') = 1$ when neither B nor ∂B contains u. If $u \in B$, then $\rho(U_B\sigma, U_B\sigma') = 0$.

Since the block dynamics determined by $\mathcal{B}$ is assumed to be contracting, there is a constant γ such that

$$\begin{aligned}\gamma\ell^d/N^d \le \mathbf{E}\Delta &= \mathbf{P}(B \ni u) - \frac{1}{N^d}\sum_{\partial B\ni u} \rho(U_B\sigma, U_B\sigma') \\ &= \frac{\ell^d}{N^d} - \frac{1}{N^d}\sum_{\partial B\ni u} \rho(U_B\sigma, U_B\sigma'),\end{aligned} \tag{4}$$

where Δ is the decrease in average Kantorovich distance between σ and σ' caused by the update.

Suppose now that we choose $\vec{j} = (j_1, \ldots, j_d)$ uniformly at random in $\{0, \ldots, \ell\}^d$ and update (in the normal fashion) all the blocks $B_{\vec{j}+(\ell+1)\vec{k}}$, where $\vec{k} \in [0, \frac{N}{\ell+1})^d$. These blocks are disjoint, and, moreover, no block has an exterior neighbor belonging to another block, hence it makes no difference in what order the updates are made. We call this series of updates a "global block update," and claim that it is contracting—meaning, in this case, that a *single* global block update reduces the Kantorovich distance between any two configurations σ and τ by a constant factor $1-\gamma/3$.

To see this, we reduce to the case where σ and $\tau = \sigma'$ differ only at a vertex u and average over choice of $\vec{j}$ to get that the expected decrease in Kantorovich distance is

$$\frac{\ell^d}{(\ell+1)^d} - \frac{1}{(\ell+1)^d} \sum_{\partial B \ni u} \rho(U_B\sigma, U_B\sigma')),$$

which, by comparing with (4), exceeds $\gamma\,(\ell/(\ell+1))^d \geq \gamma/3$ since $\ell > d$.

For $\delta = \delta(\gamma) > 0$ that we will specify later, suppose that $t_1 = t_1(\delta)$ has the following property: for any $t > t_1$, performing t single-site updates uniformly at random on the sites *inside* a block B, suffices (regardless of boundary spins) to bring the Kantorovich distance between the resulting configuration on B and the block-update configuration down to at most δ. (In fact, we can take $t_1 = 2t_0 \log(\ell^d/\delta)$.) Letting $W_B^t\sigma$ denote the distribution that results when t single-site updates are performed on B, the triangle inequality gives

$$\rho(W_B^t\sigma, W_B^t\sigma') \leq \rho(U_B\sigma, U_B\sigma') + 2\delta\,,$$

for all $t > t_1$.

Suppose next that $\mathbf{T}$ is a nonnegative-integer-valued random variable that satisfies $\mathbf{P}(\mathbf{T} < t) < \delta/\ell^d$. Since the Hamming distance of any two configurations is bounded by ℓ^d, if we perform $\mathbf{T}$ random single-site updates on the block B, we get

$$\rho(W_B^{\mathbf{T}}\sigma, W_B^{\mathbf{T}}\sigma') \leq \rho(W_B^t\sigma, W_B^t\sigma') + \ell^d\mathbf{P}(\mathbf{T} < t) \leq \rho(U_B\sigma, U_B\sigma') + 3\delta. \qquad (5)$$

Suppose we select, uniformly at random, $2tN^d/\ell^d$ sites in V. For any block B, the number of times that we select a site from B will be a binomially distributed random variable $\mathbf{T}$ with mean $2t$; its probability of falling below t is bounded above by $e^{-t/4}$ (see, e.g., [1], Thm. A.1.13, p. 312). By taking $t \geq \max\{t_1, 4\log(\ell^d/\delta)\}$ we ensure that $\mathbf{P}(\mathbf{T} < t) \leq \delta/\ell^d$ as required for (5). Note that t depends only on ℓ, d, t_1 and γ.

Let W denote the following global update procedure: choose $\vec{j}$ uniformly at random as above, perform $2tN^d/\ell^d$ random single site updates, but *censor* all updates of sites not in $\bigcup_{\vec{k}} B_{\vec{j}+(\ell+1)\vec{k}}$. To bound the expected distance between $W\sigma$ and $W\sigma'$, it suffices to consider blocks B such that u is in B or in the exterior boundary ∂B. Only one block in our global block update can contain u, and the expected number of blocks with $u \in \partial B$ is $2d/(\ell+1) < 2$, using our assumption that $\ell > d$. Therefore

$$\mathbf{E}\rho(W\sigma, W\sigma') \leq 1 - \gamma/3 + (1 + \frac{2d}{\ell+1})3\delta \leq 1 - \gamma/3 + 9\delta.$$

Taking $\delta = \gamma/36$, the RHS is at most $1 - \gamma/12$. Thus for any two configurations σ, τ, the censored Glauber dynamics procedure above yields

$$\mathbf{E}\rho(W\sigma, W\tau) \leq (1 - \frac{\gamma}{12})H(\sigma, \tau).$$

We deduce that $O(\log |V|)$ iterations of W suffice to reduce the maximal Kantorovich distance from its initial value $|V|$ (the Hamming distance between the top and bottom configurations) to any desired small constant. Recall that Kantorovich distance dominates total variation distance, and each application of W involves $2tN^d/\ell^d = O(|V|)$ single site updates, with censoring of updates that fall on the (random) boundary. Thus with this censoring, uniformly random single-site updates mix in time $\mathrm{O}(|V| \log |V|)$.

By Theorem 1.1, censoring these updates cannot improve mixing time, hence the mixing time for standard single-site Glauber dynamics is again $\mathrm{O}(|V| \log |V|)$. □

In the above theorem the periodic boundary and divisibility condition were assumed only for convenience in the proof, variations of which can be applied in many other settings. Indeed, since we announced our censoring inequality in 2001, other applications to block dynamics have been made by Martinelli and Sinclair [18], Martinelli and Toninelli [20], Mossel and Sly [21], Ding, Lubetzky and Peres [5], and Ding and Peres [6].

Note that even if the Markov random field is not monotone, our proof shows mixing time $\mathrm{O}(|V| \log |V|)$ for censored single-site dynamics; this improves by a log factor Corollary 3.3 of Van den Berg and Brouwer [2].

2. Proof of the Censoring Inequality (Theorem 1.1)

Lemma 2.1. *Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system, let μ be any distribution on Ω, and let μ_v be the result of updating μ at the site $v \in V$. If μ/π is increasing on Ω, then so is μ_v/π.*

Proof. Define $f : S^V \to \mathbb{R}$ by

$$f(\sigma) := \max \left\{ \frac{\mu(\omega)}{\pi(\omega)} \ : \ \omega \in \Omega, \ \ \omega \leq \sigma \right\} \tag{6}$$

with the convention that $f(\sigma) = 0$ if there is no $\omega \in \Omega$ satisfying $\omega \leq \sigma$. Then f is increasing on S^V, and f agrees with μ/π on Ω.

Let $\sigma < \tau$ be two configurations in Ω; we wish to show that

$$\frac{\mu_v}{\pi}(\sigma) \leq \frac{\mu_v}{\pi}(\tau). \tag{7}$$

Note first that for any $s \in S$,

$$f(\sigma_v^s) \leq f(\tau_v^s)\,,$$

since f is increasing. Furthermore, $f(\tau_v^s)$ is an increasing function of s. Thus, by (1),

$$\begin{aligned}\frac{\mu_v}{\pi}(\sigma) &= \frac{\mu(\sigma_v^\bullet)}{\pi(\sigma_v^\bullet)} = \sum_{s \in S} f(\sigma_v^s) \frac{\pi(\sigma_v^s)}{\pi(\sigma_v^\bullet)} \\ &\leq \sum_{s \in S} f(\tau_v^s) \frac{\pi(\sigma_v^s)}{\pi(\sigma_v^\bullet)} \leq \sum_{s \in S} f(\tau_v^s) \frac{\pi(\tau_v^s)}{\pi(\tau_v^\bullet)} = \frac{\mu_v}{\pi}(\tau)\,,\end{aligned}$$

where the last inequality follows from the stochastic domination guaranteed by monotonicity of the system. □

Lemma 2.2. *Suppose that S is totally ordered. If α and β are probability distributions on S such that α/β is increasing on S and $\beta(s) > 0$ for all $s \in S$, then $\alpha \succeq \beta$.*

Proof. Let g be any increasing function on S; then, with all sums taken over $s \in S$,

$$\sum g(s)\alpha(s) = \sum g(s)\frac{\alpha(s)}{\beta(s)}\beta(s) \geq \sum g(s)\beta(s) \cdot \sum \frac{\alpha(s)}{\beta(s)}\beta(s) = \sum g(s)\beta(s),$$

confirming stochastic domination. The inequality in the chain is the positive correlations property of totally ordered sets (which goes back to Chebyshev, see [14] §II.2), applied to the increasing functions g and α/β on S with measure β. □

Lemma 2.3. *Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system. If μ is a distribution on Ω such that μ/π is increasing, then $\mu \succeq \mu_v$ for any $v \in V$.*

Proof. Let g be increasing. If $\sigma \in \Omega$ satisfies $\mu(\sigma_v^\bullet) > 0$, then μ/μ_v is increasing on $\sigma_v^\bullet$. By Lemma 2.2 (applied to $\{s \in S : \sigma_v^s \in \Omega\}$ in place of S), for such σ we have

$$\sum_{s \in S} g(\sigma_v^s)\frac{\mu(\sigma_v^s)}{\mu(\sigma_v^\bullet)} \geq \sum_{s \in S} g(\sigma_v^s)\frac{\mu_v(\sigma_v^s)}{\mu(\sigma_v^\bullet)} \,.$$

Multiplying by $\mu(\sigma_v^\bullet)$ and summing over all choices of $\sigma_v^\bullet$ gives

$$\sum_{\sigma \in \Omega} g(\sigma)\mu(\sigma) \geq \sum_{\sigma \in \Omega} g(\sigma)\mu_v(\sigma)\,,$$

establishing the required stochastic dominance. □

Lemma 2.4. *Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system, and let μ, ν be two arbitrary distributions on Ω. If ν/π is increasing on Ω and $\nu \preceq \mu$, then $\|\nu - \pi\| \leq \|\mu - \pi\|$.*

Proof. Let $A = \{\sigma : \nu(\sigma) > \pi(\sigma)\}$. Then the indicator of A is increasing, so

$$\|\nu - \pi\| = \sum_{\sigma \in A}(\nu(\sigma) - \pi(\sigma)) = \nu(A) - \pi(A) \leq \mu(A) - \pi(A),$$

since $\nu \preceq \mu$. The right-hand side is at most $\|\mu - \pi\|$. □

Theorem 2.5. *Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system. Let μ be the distribution on Ω which results from successive updates at sites $u_1, \ldots, u_k$, beginning at the top configuration. Let ν be defined similarly but with the update at u_j left out. Then*

1. *$\mu \preceq \nu$, and*
2. *$\|\mu - \pi\| \leq \|\nu - \pi\|$.*

Proof. Let μ^0 be the distribution concentrated at the top configuration, and $\mu^i = (\mu^{i-1})_{u_i}$ for $i \geq 1$. Applying Lemma 2.1 inductively, we have that each μ^i/π is increasing, for $0 \leq i \leq k$. In particular, we see from Lemma 2.3 that $\mu^{j-1} \succeq (\mu^{j-1})_{u_j} = \mu^j$.

If we define ν^i in the same manner as μ^i, except that $\nu^j = \nu^{j-1}$, then because stochastic dominance persists under updates, we have $\nu^i \succeq \mu^i$ for all i; when $i = k$, we get $\mu \preceq \nu$ as desired.

For the second statement of the theorem, we merely apply Lemma 2.4, noting that ν^k/π is increasing by the same inductive argument used for μ. □

Proof of Theorem 1.1. Apply Theorem 2.5 inductively, censoring one site at a time. This establishes the case where the update locations are deterministic. In the case where the update sequence $v_1(\xi) \ldots, v_m(\xi)$ that yields μ is random (defined on some probability space $(\Xi, \mathbf{P}_\Xi)$) and its subsequence leading to ν is also random (defined on the same probability space), then conditioning on ξ yields measures $\mu(\xi)$ and $\nu(\xi)$ such that $\mu(\xi) \preceq \nu(\xi)$ and $\nu(\xi)/\pi$ is increasing on Ω. These properties are preserved under averaging over Ξ, so we conclude that $\mu \preceq \nu$ and ν/π is increasing on Ω. The inequality between total variation norms follows from Lemma 2.4. □

Corollary 2.6. *Let $\langle \Omega, S, V, \pi \rangle$ be the ferromagnetic Ising model on a graph G (with arbitrary boundary conditions and external field). Let μ be the distribution on Ω which results from successive block updates at vertex sets $B_1, \ldots, B_k$, beginning at the top configuration. Let ν be defined similarly but with the update at B_j left out. Then*

1. *$\mu \preceq \nu$, and*
2. *$\|\mu - \pi\| \leq \|\nu - \pi\|$.*

Proof. Each block update can be approximated within ϵ (in total variation) by running single-site dynamics in the block for sufficiently long. Applying Theorem 2.5 and letting $\epsilon \to 0$ completes the proof. □

3. Comparison of Single Site Update Schemes

In practice, updates on a system $\langle \Omega, S, V, \pi \rangle$ are often performed systematically rather than at random. Typically a permutation of V is fixed and sites are updated periodically in permutation order. If the interaction graph is bipartite, it is possible and often convenient to update all odd sites simultaneously, then all even sites, and repeat; we call this *alternating* updates. To be fair, we count a full round of alternating updates as n single updates, so that alternating updates constitute a special case of systematic updating.

Mixing time may differ from one update scheme to another; for example, if there are no interactions (so that one update per site produces perfect mixing) then systematic updating is faster by a factor of $\frac{1}{2} \log n$ than uniformly random updates, since after $(\frac{1}{2} - \varepsilon)n \log n$ random updates about $n^{1/2+\varepsilon}$ sites have not been hit, so counting the number of sites that still have the initial spin implies the total variation distance to equilibrium is still close to 1. (For a more general $\Omega(n \log n)$ lower bound for Glauber dynamics with random updates see [10]).

Embarrassingly, there are only a few results to support the observation that mixing times for the various update schemes never seem to differ by more than a factor of $\log n$ and rarely by more than a constant. (See [7,8] for some recent progress in the Dobrushin uniqueness regime.) Theorem 1.1 allows us to obtain some useful comparison results for monotone systems, but is still well short of what is suspected to be true.

Theorem 3.1. *Let $\mathcal{A}$ be the alternating update scheme, and $\mathcal{S}$ an arbitrary systematic update scheme, for a bipartite monotone system $\langle \Omega, S, V, \pi \rangle$. Then the mixing time for $\mathcal{S}$ (starting at the top state) is no more than twice the mixing time for $\mathcal{A}$.*

Proof. When updating according to $\mathcal{S}$, we censor all even-site updates; on even passes, all odd-site updates. Since successive updates of sites of the same parity commute, the result is exactly $\mathcal{A}$ and an application of Theorem 1.1 shows that we mix at a cost of at most a factor of 2. □

Theorem 3.2. *Let $\mathcal{A}$ be the alternating update scheme, and $\mathcal{R}$ the uniformly random update scheme, for a bipartite monotone system $\langle \Omega, S, V, \pi \rangle$. Then the mixing time for $\mathcal{R}$ (starting at the top state) is no more than $2 \log n$ times the mixing time for $\mathcal{A}$.*

Proof. When updating according to $\mathcal{R}$, we censor all even-site updates until all odd sites are hit; then we censor all odd-site updates until all even sites are hit, and repeat. Since each of these steps takes $2(n/2)\log(n/2)$ updates on average, Theorem 1.1 guarantees a loss of at most a factor of $2\log n$. □

Theorem 3.3. *Let $\mathcal{R}$ be the uniformly random update scheme, and $\mathcal{S}$ an arbitrary systematic update scheme, for a monotone system $\langle \Omega, S, V, \pi \rangle$ of maximum degree $\Delta_{\max}$. Then the mixing time for $\mathcal{S}$ is no more than $O(\sqrt{\Delta_{\max} n})$ times the mixing time for $\mathcal{R}$.*

Proof. Prior to implementing a round of $\mathcal{S}$, we choose uniformly random sites one by one as long as no two are adjacent; since the probability of adjacency for a random pair of sites is at most $(\Delta_{\max} + 1)/n$, this "birthday problem" procedure will keep about $\sqrt{n/\Delta_{\max}}$ updates. All updates of sites not on this list are censored from the upcoming round of $\mathcal{S}$, incurring a loss of a factor of $n/\sqrt{n/\Delta_{\max}} = \sqrt{\Delta_{\max} n}$. Since updates of non-adjacent sites commute, Theorem 1.1 applies. □

If $\langle \Omega, S, V, \pi \rangle$ is bipartite, then since the alternating scheme is a systematic scheme, Theorem 3.3 applies to it as well.

From systematic updates to alternating or random updates, there seems to be nothing better to do in our context than to score one update per systematic round, incurring a factor of n penalty.

3.1. Hanging Subgraphs. Let H be a subgraph of the finite graph G, on which some system $\langle \Omega, S, V, \pi \rangle$ is defined, and suppose what is wanted is mixing on H. When continuous-time Glauber dynamics is employed, it is natural to compare mixing time $T_H(\epsilon)$ on H by itself (that is, with the rest of G destroyed) with the mixing time $T_G(\epsilon)$ when all points of G are being updated. Indeed, consider the Ising model (with no external field and free boundary conditions), we conjecture that $T_H(\epsilon)$ never exceeds $T_G(\epsilon)$—echoing a conjecture of the first author for spectral gaps, cited in [22] and proved there when G is a cycle.

Because the Ising model is a Markov random field, and its stationary distribution on a single site is independent of the graph (since we assumed a free boundary and no external field), it enjoys the following property: if only one vertex (say, x) of H is adjacent to vertices of $G\backslash H$, then the stationary distribution on H is identical to the stationary distribution on G restricted to H. To see this, it suffices to note that either stationary distribution can be obtained by flipping a coin to determine the sign of x, then conditioning the rest of the configuration on the result.

We can now make use of Theorem 1.1, together with monotonicity of the Ising model, to prove our conjecture in this limited case.

Theorem 3.4. *Let H be a subgraph of the finite graph G and suppose that at most one vertex of H is adjacent to vertices of $G\backslash H$. Begin in the all "+" state and fix a mixing tolerance ϵ for continuous Glauber dynamics. Then $T_H(\epsilon, +) \le T_G(\epsilon, +)$.*

Proof. The result is of course trivial if H is disconnected from $G\backslash H$; otherwise let x be the unique vertex of H with neighbors outside H. Run continuous-time Glauber dynamics on G for t time units and let $Q = \langle v_1, \ldots, v_k \rangle$ be the random sites updated. Let Q' on G be the result of replacing each update of x in Q by a block update of $\{x\} \cup (G\backslash H)$. Then, on account of the property noted above, the effect of the Q' update on H is identical to running continuous time Glauber dynamics on H. The theorem now follows from repeated application of Corollary 2.6. □

Acknowledgements. We are grateful to Dana Randall for telling us about the censoring idea of Van den Berg and Brouwer and for suggesting the question addressed in this paper. We also thank the participants of the 2001 IES in Sweden, especially Svante Janson and Russell Lyons, for helpful discussions during the early stages of this work. Finally, we thank Subhro Ghosh and two referees for useful comments.

References

1. Alon, N., Spencer, J.: *The Probabilistic Method*, 3^{rd} ed., Hoboken, New Jersey: John Wiley & Sons, 2008
2. van den Berg, J., Brouwer, R.: Random sampling for the monomer-dimer model on a lattice. J. Math. Phys. **41**, 1585–1597 (1999)
3. Berger, N., Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs. Prob. Th. Rel. Fields **131**, 311–340 (2005)
4. Bubley, R., Dyer, M.: Path coupling: a technique for proving rapid mixing in Markov chains. In: *Proceedings of the* 38^{th} *Annual Symposium on Foundations of Computer Science (FOCS)*, Los Alamitos, CA: IEEE Comp. Soc., 1997, pp. 223–231
5. Ding, J., Lubetzky, E., Peres, Y.: Mixing time of critical Ising model on trees is polynomial in the height. Commun. Math. Phys. **295**, 161–207 (2010)
6. Ding, J., Peres, Y.: Mixing time for the Ising model: a uniform lower bound for all graphs. Ann. l'Inst. H. Poincaré - Prob. Et Stat. **47**, 1020–1028 (2011)
7. Dyer, M., Goldberg, L.A., Jerrum, M.: Systematic scan for sampling colourings. Ann. Appl. Prob. **16**, 185–230 (2006)
8. Dyer, M., Goldberg, L.A., Jerrum, M.: Dobrushin conditions and systematic scan. Comb. Prob. Comp. **17**, 761–779 (2008)
9. Dyer, M., Sinclair, A., Vigoda, E., Weitz, D.: Mixing in time and space for lattice spin systems: A combinatorial view. Rand. Struct. Alg. **24**, 461–479 (2004)
10. Hayes, T.P., Sinclair, A.: A general lower bound for mixing of single-site dynamics on graphs. In: 46^{th} *Annual IEEE Symposium on Foundations of Computer Science* (FOCS 2005), Los Alamitos, CA: IEEE Comp. Soc., 2005, pp. 511–520
11. Holroyd, A.: Some circumstances where extra updates can delay mixing. J. Stat. Phys. **145**, 1649–1652 (2011)
12. Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs (extended abstract). In: 42^{nd} *IEEE Symposium on Foundations of Computer Science (FOCS), Las Vegas, NV, 2001)*, *Los Alamitos, CA: IEEE Computer Soc.*, 2001, pp. 568–578
13. Levin, D. A., Peres, Y., Wilmer, E.: *Markov Chains and Mixing Times*, Providence, RI: Amer. Math. Soc., 2008
14. Liggett, T.: *Interacting particle systems*. New York: Springer, 1985
15. Martinelli, F.: Lectures on Glauber dynamics for discrete spin models. In: *Lectures on probability theory and statistics (Saint-Flour, 1997) Lecture Notes in Math.* **1717**, Berlin: Springer, 1998, pp. 93–191
16. Martinelli, F.: Relaxation times of Markov chains in statistical mechanics and combinatorial structures. In: *Encyclopedia of Mathematical Sciences*, Vol. **110**, Berlin-Heidelberg-New York: Springer, 2003, pp. 175–272
17. Martinelli, F., Olivieri, E.: Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case. Commun. Math. Phys. **161**, 447–486 (1994)
18. Martinelli, F., Sinclair, A.: Mixing time for the Solid-on-Solid model. In: *Proc. ACM STOC* 2009, New York: ACM, 2009, pp. 571–580
19. Martinelli, F., Sinclair, A., Weitz, D.: Glauber dynamics on trees: Boundary conditions and mixing time. Commun. Math. Phys. **250**, 301–334 (2004)
20. Martinelli, F., Toninelli, F.: On the mixing time of the 2D stochastic Ising model with "plus" boundary conditions at low temperature. Commun. Math. Phys. **296**, 175–213 (2010)
21. Mossel, E., Sly, A.: Exact Thresholds for Ising-Gibbs Samplers on General Graphs. Ann. Prob. **41**(1), 294–328 (2013)
22. Nacu, S.: Glauber dynamics on the cycle is monotone. Prob. Th. Rel. Fields **127**, 177–185 (2003)
23. Peres, Y.: Lectures on Mixing for Markov Chains and Spin Systems. University of British Columbia, Summary at http://www.stat.berkeley.edu/~peres/ubc.pdf, 2005

Communicated by F. L. Toninelli