Can extra updates delay mixing?

Yuval Peres ∗ Peter Winkler †

July 7, 2010

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 and the total variation distance to the stationary distribution. We deduce that for monotone Markov random fields, when block dynamics contracts a weighted Hamming metric, so does single-site dynamics started at extremal configurations. 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 to compare block and single-site dynamics; it has already been used for this purpose in works of Martinelli-Sinclair, Mossel-Sly, Ding, Lubetzky and the authors.

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 to transport bounds for block dynamics to bounds for single site dynamics; sharp estimates result in several situations.

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

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

∗Microsoft Research, Redmond
†Bradley Third Century Professor in the Sciences, Dartmouth College.
where $Z$ is the appropriate normalizing constant. More generally, our results apply when $\pi$ 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 $\pi$ conditioned on the spins of its neighbors. The resulting Markov chain is irreducible, aperiodic and has unique stationary distribution $\pi$. Let $p^t(\omega, \cdot)$ be the distribution of configurations after $t$ steps, with initial state $\omega$. Let $\|\mu - \nu\| = \frac{1}{2} \sum_{\sigma} |\mu(\sigma) - \nu(\sigma)|$ be the total variation distance. The mixing time $T_G(\epsilon)$ for the dynamics is the least $t$ such that $\|p^t(\omega, \cdot) - \pi\| \leq \epsilon$ for any $\omega \in \Omega$. In discrete-time block dynamics, a family $B$ of “blocks” of sites is provided. At each step, a block $B \in B$ is selected uniformly at random and a configuration on $B$ is selected according to $\pi$ 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 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, 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 $O(n \log n)$ steps in total variation.

- 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 in total variation. (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 $G \setminus H$,
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 at most by a factor of \( \log n \) than uniform random updates.

See §1.2 for further discussion of block dynamics, and §2-3 for proofs.

### 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 \( \pi \) on \( \Omega \), 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 \( \Omega \) 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 \( \sigma^s_v \) the configuration obtained from \( \sigma \) by changing its value at \( v \) to \( s \), that is, \( \sigma^s_v(v) = s \) and \( \sigma^s_v(u) = \sigma(u) \) for all \( u \neq v \). Let \( \sigma^*_v \) be the set of configurations \( \{\sigma^s_v\}_{s \in S} \) in \( \Omega \).

The update \( \mu_v \) at \( v \) of a distribution \( \mu \) on \( \Omega \) is defined by

\[
\mu_v(\sigma) = \frac{\pi(\sigma)}{\pi(\sigma^*_v)} \mu(\sigma^*_v) \quad \text{for } \sigma \in \Omega. \tag{1}
\]

For measures \( \mu \) and \( \nu \) on a poset \( \Gamma \), we write \( \nu \preceq \mu \) to indicate that \( \mu \) stochastically dominates \( \nu \), 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

\[
\left\{ \frac{\pi(\sigma^s_v)}{\pi(\sigma^*_v)} \right\}_{s \in S} \preceq \left\{ \frac{\pi(\tau^s_v)}{\pi(\tau^*_v)} \right\}_{s \in S} \tag{2}
\]

as distributions on the spin set \( S \).
1.2 Main results

**Theorem 1.1.** Let \( \langle \Omega, S, V, \pi \rangle \) be a monotone system and let \( \mu \) be the distribution on \( \Omega \) which results from successive updates at sites \( v_1, \ldots, v_m \), beginning at the top configuration. Let \( \nu \) be defined similarly but with updates only at a subsequence \( v_{i_1}, \ldots, v_{i_k} \). Then \( \mu \preceq \nu \), and \( \| \mu - \pi \| \leq \| \nu - \pi \| \) in total variation. Moreover, this also holds if the sequence \( v_1, \ldots, v_m \) and the subsequence \( i_1, \ldots, i_k \) are chosen at random according to any prescribed distribution. See Section 2 for the proof. The proof shows that the assumption of starting from the top configuration can be replaced by the assumption that the dynamics starts at a distribution \( \mu_0 \) where the likelihood ratio \( \mu_0/\pi \) is weakly increasing.

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

Suppose Endow \( \Omega \subset S^V \) with the \( \ell^1 \) metric

\[
\rho(\sigma, \tau) = \sum_{v \in V} |\sigma_v - \tau_v|.
\]

(More generally, it is sometimes fruitful to consider a weighted \( \ell^1 \) metric). Given a subset \( B \) of \( V \), let \( \sigma_B^* \) be the set of configurations \( \tau \in \Omega \) such that \( \tau \) agrees with \( \sigma \) off \( B \). For \( \sigma \in \Omega \), the block update \( U_B \sigma \) is a measure on \( \sigma_B^* \) defined by \( (U_B \sigma)(\omega) = \frac{\pi(\omega)}{\pi(\sigma_B^*)} \) for \( \omega \in \sigma_B^* \). Thus \( U_B \sigma \) is \( \pi \) conditioned on \( \sigma_B^* \). For a collection of blocks \( B \), the \( B \)-averaged block update of \( \sigma \in \Omega \) yields a random configuration with distribution \( \frac{1}{|B|} \sum_{B \in B} U_B \sigma \). We call the block dynamics contracting if any two configurations \( \sigma \leq \tau \), the expected number of discrepancies after a block update is smaller by a factor \( 1 - \gamma/|V| \).

where \( \gamma \) is a constant. In our setting, contraction implies a bound of order \( |V| \log |V| \) on the mixing time, since the number of blocks is of order \( |V| \).

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 \( \sigma \) to \( \partial B \), the set of vertices in \( B^c \) that are adjacent to \( B \).

Recall that in discrete time (heat-bath) Glauber dynamics, at each step a site in \( V \) is chosen at random to update.

For the next result, we will assume that \( S \subset \mathbb{R} \) with the induced total order. For simplicity,
we consider $V$ to be a box in the lattice $\mathbb{Z}^d$ and the blocks will be sub-boxes of bounded size. However, our method is applicable in a variety of other settings.

**Theorem 1.2.** Consider a monotone Markov random field on the box $V = [0, N]^d$. For each cube of side length $b$ that intersects $V$ define a block by intersecting that box with $V$. If the corresponding block dynamics is contracting and the single-site dynamics restricted to a block has a bounded mixing time (uniformly in boundary conditions), then the corresponding single-site dynamics is also contracting, so it has mixing time $O(|V| \log |V|)$.

See [23] for an exposition of this result which is due to both authors of the present note (It will be presented in greater detail in the full version.)

Since we discovered our censoring inequality in 2001, Other applications to block dynamics were made by Martinelli and Sinclair [18], Martinelli and Toninelli [20], Mossel and Sly [21], Ding, Lubetzky and Peres [5] and Ding-Peres [6]. In particular, the latter paper uses the censoring inequality to prove a uniform lower bound asymptotic to $n \log(n)/4$ for the mixing time of Glauber dynamics of the Ising model on any $n$-vertex graph.

## 2 Proof of the censoring inequality

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

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

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

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 $\mu/\pi$ on $\Omega$.

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

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

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

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

$$
\frac{\mu_v}{\pi}(\sigma) = \frac{\mu(\sigma_v^s)}{\pi(\sigma_v^s)} = \sum_{s \in S} f(\sigma_v^s) \frac{\pi(\sigma_v^s)}{\pi(\sigma_v^s)}
$$

$$
\leq \sum_{s \in S} f(\tau_v^s) \frac{\pi(\sigma_v^s)}{\pi(\sigma_v^s)} \leq \sum_{s \in S} f(\tau_v^s) \frac{\pi(\tau_v^s)}{\pi(\tau_v^s)} = \mu_v(\tau),
$$

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

\begin{lemma}
Suppose that $S$ is totally ordered. If $\alpha$ and $\beta$ are probability distributions on $S$ such that $\alpha/\beta$ is increasing on $S$ and $\beta(s) > 0$ for all $s \in S$, then $\alpha \succeq \beta$.
\end{lemma}

\begin{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 [13] §II.2), applied to the increasing functions $g$ and $\alpha/\beta$ on $S$ with measure $\beta$. 

\end{proof}

\begin{lemma}
Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system. If $\mu$ is a distribution on $\Omega$ such that $\mu/\pi$ is increasing, then $\mu \succeq \mu_v$ for any $v \in V$.
\end{lemma}

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

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

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

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

establishing the required stochastic dominance. 

\end{proof}

\begin{lemma}
Let $\langle \Omega, S, V, \pi \rangle$ be a monotone system, and let $\mu, \nu$ be two arbitrary distributions on $\Omega$. If $\nu/\pi$ is increasing on $\Omega$ and $\nu \succeq \mu$, then $\|\nu - \pi\| \leq \|\mu - \pi\|$.
\end{lemma}
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 \( \| \mu - \pi \| \).

\[ \square \]

**Theorem 2.5.** Let \( \langle \Omega, S, V, \pi \rangle \) be a monotone system, Let \( \mu \) be the distribution on \( \Omega \) which results from successive updates at sites \( u_1, \ldots, u_k \), beginning at the top configuration. Let \( \nu \) 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 \( \mu^0 \) 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 \( \mu^i/\pi \) 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 \( \nu^j \) in the same manner as \( \mu_i \), 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 \( \nu^k/\pi \) is increasing by the same inductive argument used for \( \mu \).

\[ \square \]

**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 \( \mu \) is random (defined on some probability space \( (\Xi, P_\Xi) \)) and its subsequence leading to \( \nu \) is also random (defined on the same probability space), then conditioning on \( \xi \) yields measures \( \mu(\xi) \) and \( \nu(\xi) \) such that \( \mu(\xi) \preceq \nu(\xi) \) and \( \nu(\xi)/\pi \) is increasing on \( \Omega \). These properties are preserved under averaging over \( \Xi \), so we conclude that \( \mu \preceq \nu \) and \( \nu/\pi \) is increasing on \( \Omega \). The inequality between total variation norms follows from Lemma 2.4.

\[ \square \]

### 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 permuta-
tion 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 $(\frac{1}{2} - \epsilon)n\log n$ random updates about $n^{1/2+\epsilon}$ 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 [11]).

Embarrassingly, there are 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 $A$ be the alternating update scheme, and $S$ an arbitrary systematic update scheme, for a bipartite monotone system $\langle \Omega, S, V, \pi \rangle$. Then the mixing time for $S$ is no more than twice the mixing time for $A$.

**Proof.** When updating according to $S$, 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 $A$ 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 $A$ be the alternating update scheme, and $R$ the uniformly random update scheme, for a bipartite monotone system $\langle \Omega, S, V, \pi \rangle$. Then the mixing time for $R$ is no more than $\log n$ times the mixing time for $A$.

**Proof.** When updating according to $R$, 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 the natural log of $n$. □
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 $\Delta$. Then the mixing time for $\mathcal{S}$ is no more than $O(\sqrt{\Delta n})$ 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 $(\Delta + 1)/n$, this “birthday problem” procedure will keep about $\sqrt{(n/\Delta)}$ 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 $n/(\sqrt{(n/\Delta)} = \sqrt{\Delta n}$. 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 are employed, it is natural to compare mixing time $T_H$ on $H$ by itself (that is, with the rest of $G$ destroyed) with mixing time $T_G$ when all points of $G$ are being updated. Indeed, for the Ising model (with no external field), we conjecture that $T_H$ never exceeds $T_{G|H}$—echoing a conjecture of the first author for spectral gaps, cited in [22] and proved there when $G$ is a cycle. Putting it another way, we think bigger is slower.

Because the Ising model is a Markov random field, and its stationary distribution on a single site is independent of the graph, it enjoys the following property: if only one vertex (say, $x$) of $H$ is adjacent to vertices of $G \setminus H$, 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 \( G \setminus H \). Begin in the all “+” state and fix a mixing tolerance \( \epsilon \) for continuous Glauber dynamics. Then \( T_H(\epsilon) \leq T_{G \setminus H}(\epsilon) \).

**Proof.** The result is of course trivial if \( H \) is disconnected from \( G \setminus H \); otherwise let \( x \) be the unique vertex of \( H \) with neighbors outside \( H \). Let \( Q = \{v_1, \ldots, v_k\} \) be the target sites of a sequence of updates on \( H \), and let \( Q' \) on \( G \) be the result of replacing each update of \( x \) in \( Q \) by a block update of \( \{x\} \cup (G \setminus H) \). Then, on account of the property noted above, the effects of \( Q \) and \( Q' \) are identical on \( H \).

If it were not the case that \( T_H(\epsilon) \leq T_{G \setminus H}(\epsilon) \), then there would in particular be an update sequence \( Q \) for \( H \) and a supersequence \( Q^+ \) for \( G \), all added sites being outside \( H \), such that \( Q^+ \) gets \( H \) closer by some \( \delta > 0 \) to stationarity than does \( Q \). However, for large enough \( j \), we can replace the block updates in \( Q' \) by \( j \) single-site updates within \( \{x\} \cup (G \setminus H) \) to get a new update sequence \( Q'' \) which contains \( Q^+ \), but whose resulting distribution matches that of \( Q' \) (thus also \( Q \)) to within total variation \( \delta/2 \). This would force \( Q'' \) to mix better than \( Q^+ \), contradicting Theorem 1.1. \( \square \)

**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.

**References**


