Entropy decay in the Swendsen-Wang dynamics

Antonio Blanca∗ Pietro Caputo† Daniel Parisi‡ Alistair Sinclair§ Eric Vigoda¶

July 15, 2020

Abstract

We study the mixing time of the Swendsen-Wang dynamics for the ferromagnetic Ising and Potts models on the integer lattice \( \mathbb{Z}^d \). This dynamics is a widely used Markov chain that has largely resisted sharp analysis because it is non-local, i.e., it changes the entire configuration in one step. We prove that, whenever strong spatial mixing (SSM) holds, the mixing time on any \( n \)-vertex cube of \( \mathbb{Z}^d \) is \( O(\log n) \), improving on the previous best known bound of \( O(n) \). SSM is a standard condition corresponding to exponential decay of correlations with distance between spins on the lattice and is known to hold in \( d = 2 \) dimensions throughout the high-temperature (single phase) region. Our result follows from a modified log-Sobolev inequality, which expresses the fact that the dynamics contracts relative entropy at a constant rate at each step. The proof of this fact utilizes a new factorization of the entropy in the joint probability space over spins and edges that underlies the Swendsen-Wang dynamics. This factorization leads to several additional results, including mixing time bounds for a number of natural local and non-local Markov chains on the joint space, as well as for the standard random-cluster dynamics.

∗Department of Computer Science and Engineering, Penn State, University Park, PA 16801. Email: ablanca@cse.psu.edu. Research supported in part by NSF grant CCF-1850443.
†Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy. Email: pietro.caputo@uniroma.3.it
‡Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy. Email: daniel.parisi@uniroma3.it
§Computer Science Division, U.C. Berkeley, Berkeley, CA 94720. Email: sinclair@cs.berkeley.edu. Research supported in part by NSF grant CCF-1815328.
¶School of Computer Science, Georgia Tech, Atlanta, GA 30332. Email: vigoda@gatech.edu. Research supported in part by NSF grant CCF-1563838.
1 Introduction

The ferromagnetic Potts model is a classical spin system in statistical physics. It is the generalization of the Ising model from two to many spins, and has found applications in machine learning, computer vision, computational biology and social networks. The Potts model is specified by a graph $G = (V,E)$, a set of spins (or colors) $[q] = \{1, \ldots, q\}$, and an edge weight or inverse temperature parameter $\beta > 0$. A configuration $\sigma \in \Omega = \{1, \ldots, q\}^V$ of the model assigns a spin value to each vertex $v \in V$, and the probability of finding the system in a given configuration $\sigma$ is given by the Gibbs (or Boltzmann) distribution

$$
\mu(\sigma) = \mu_{G,\beta}(\sigma) := \frac{1}{Z} \exp(-\beta|D(\sigma)|),
$$

where $D(\sigma) = \{\{v, w\} \in E : \sigma_v \neq \sigma_w\}$ is the set of edges whose endpoints have disagreeing spins in $\sigma$ and $Z := \sum_{\sigma \in \Omega} \exp(-\beta|D(\sigma)|)$ is the normalizing factor or partition function. Note that this model is ferromagnetic, in the sense that neighboring spins want to align with each other. The Ising model of ferromagnetism is exactly the case $q = 2$.

We focus on the classical setting where $G$ is a subgraph of the infinite $d$-dimensional lattice graph $\mathbb{Z}^d$. We will mostly restrict attention to the case of $d$-dimensional cubes, i.e., $V = \{0, \ldots, \ell\}^d$, but our results can be extended to more general geometries of $\mathbb{Z}^d$; see Remark 2.2.

A popular Markov chain for sampling from the Gibbs distribution (1.1) is the Swendsen-Wang (SW) dynamics [48], which utilizes the random-cluster representation of the Potts model to derive a sophisticated “non-local” Markov chain in which every vertex can update its spin in each step. From the current spin configuration $\sigma(t) \in \Omega$, the SW dynamics generates $\sigma(t+1) \in \Omega$ as follows:

1. Let $M(\sigma(t)) = E \setminus D(\sigma(t)) = \{\{v, w\} \in E : \sigma_v(t) = \sigma_w(t)\}$ be the set of monochromatic edges of $G$ in $\sigma(t)$.

2. Independently for each edge $e \in M(\sigma(t))$, retain $e$ with probability $1 - \exp(-\beta)$ and delete it otherwise, resulting in the subset $A(t) \subseteq M(\sigma(t))$. (This is equivalent to performing bond percolation with probability $1 - \exp(-\beta)$ on the subgraph $(V, M(\sigma(t)))$.)

3. For each connected component $C$ in the subgraph $(V, A(t))$, independently choose a spin $s_C$ uniformly at random from $[q]$ and assign $s_C$ to all vertices in $C$, yielding $\sigma(t+1) \in \Omega$.

The Swendsen-Wang dynamics is ergodic, and (1.1) is the (unique) stationary distribution of this chain; see [21] for a proof. This non-local dynamics was originally proposed as a heuristic method for overcoming the relatively slow convergence of local Markov chains close to the critical temperature, due to its ability to flip large regions of spins in one step. Though widely used in practice, and theoretically appealing, our detailed quantitative understanding of its properties is still far from complete.

In this paper, we are interested in the speed of convergence of the SW dynamics to stationarity, and in particular its mixing time. The mixing time captures the convergence rate in total variation distance of a Markov chain from the worst possible starting configuration and is the most standard measure of the speed of convergence. The mixing behavior of the SW dynamics for the Potts model is quite subtle. It is expected to converge much faster than the Glauber dynamics, a standard local Markov chain that updates the spin of a single, randomly chosen vertex at each step, yet there are multiple examples where the mixing time of the SW dynamics is exponential in the number of vertices of the graph; see, e.g., [25, 23, 8, 24, 11, 12]. Results proving tight bounds for the mixing time of the SW dynamics are rare, and are limited to very special classes of graphs, such as the
complete graph and trees [31, 34, 23, 8], or to very high temperatures [41, 45]. Most bounds for the mixing time of the SW dynamics are derived by comparison with the (much slower) Glauber dynamics [49], and are consequently often far from sharp.

There is a long line of work studying the connection between spatial mixing (i.e., decay of correlations) properties of Gibbs distributions and the speed of convergence of reversible Markov chains (see, e.g., [30, 1, 50, 47, 38, 39, 15, 20, 44]). These results focus on local Markov chains, such as the Glauber dynamics, but there has been some recent progress in understanding this connection for non-local Markov chains such as the SW dynamics [6, 5, 13]. In particular, it was established in [5] that the strong spatial mixing (SSM) property implies that the mixing time $T_{\text{mix}}(\text{SW})$ of the SW dynamics is $O(n)$, where $n := |V|$ is the number of vertices.

SSM is a standard formalization of decay of correlations in spin systems, and, roughly speaking, expresses the fact that the correlation between spins at different vertices decreases exponentially with distance between them. More precisely, given a pair of fixed configurations $\psi$ and $\psi_u$ on the boundary of $V$ such that $\psi$ and $\psi_u$ differ only in the spin of the vertex $u$, the effect on the (conditional) marginal distribution at a set $B \subseteq V$ decays exponentially with distance between $B$ and the disagreement at $u$; see Section 2 for a precise definition. Our main algorithmic result in this paper is that the mixing time of the SW dynamics is in fact $O(\log n)$ whenever SSM holds.

**Theorem 1.1.** In an $n$-vertex cube of $\mathbb{Z}^d$, for all integer $q \geq 2$, SSM implies that for all boundary conditions $T_{\text{mix}}(\text{SW}) = O(\log n)$.

We recall that a boundary condition $\tau$ for the Potts model is a fixed assignment of spins to the boundary of $V$; in the presence of a boundary condition, we consider the Gibbs distribution on $V$ conditional on the assignment $\tau$ on the boundary of $V$. The case where there is no boundary condition is known as the free boundary case and is also covered by our results. We believe that the mixing time bound in Theorem 1.1 is tight; indeed, this is known for periodic boundary conditions at sufficiently high temperatures [45].

In $\mathbb{Z}^2$, SSM is known to hold for all $q \geq 2$ and all $\beta < \beta_c(q)$, where $\beta_c(q) = \ln(1 + \sqrt{q})$ is the uniqueness threshold [4, 2, 40]. Therefore, we obtain the following immediate corollary of Theorem 1.1.

**Corollary 1.2.** In an $n$-vertex square region of $\mathbb{Z}^2$, for all $q \geq 2$, all $\beta < \beta_c(q)$ and all boundary conditions, we have $T_{\text{mix}}(\text{SW}) = O(\log n)$.

The best previous bound in the setting of Corollary 1.2 was $T_{\text{mix}}(\text{SW}) = O(n)$ and follows from the results in [5]. Nam and Sly [45] recently proved an $O(\log n)$ mixing time bound (as well as the cutoff phenomenon) for sufficiently high temperatures ($\beta \ll \beta_c(q)$) for the periodic boundary condition.

Our methods also provide new results for the low-temperature regime $\beta > \beta_c(q)$ in $\mathbb{Z}^2$ for specific boundary conditions. We say that a boundary condition $\tau$ is monochromatic if $\tau$ fixes the spin of every boundary vertex to the same color. One of the most fundamental open problems in the study of dynamics for the Ising and Potts model concerns the mixing time of the Glauber dynamics at low temperatures with a monochromatic boundary [42, 35]. We provide new bounds for the mixing time of the SW dynamics in this setting.

**Theorem 1.3.** In an $n$-vertex square region of $\mathbb{Z}^2$, for all $q \geq 2$ and all $\beta > \beta_c(q)$ we have $T_{\text{mix}}(\text{SW}) = O(n \log n)$ for the free or monochromatic boundary condition.

The best previously known bound for the mixing time of the SW dynamics in an $n$-vertex square region of $\mathbb{Z}^2$ when $\beta > \beta_c(q)$ was $O(n^2(\log n)^2)$. This bound follows from the results in [9, 49]; see
is likely not tight, and we conjecture that the SW dynamics also mixes in $O(\log n)$ steps in this regime.

The key new ingredient in proving the above results, and the main technical contribution of this paper, is the fact that, in the presence of strong spatial mixing, the SW dynamics contracts relative entropy with respect to the Gibbs distribution $\mu$ at a constant rate. This is a much stronger statement than the more standard contraction in variance, which follows from analysis of the spectral gap. To state this contraction formally, for a positive function $f : \Omega \to \mathbb{R}_+$, let $\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$ denote the mean of $f$, and $\text{Ent}_\mu(f) = \mu[f \log(f/\mu[f])]$ the entropy of $f$ with respect to $\mu$. (Note that if $f$ is normalized so that $\mu[f] = 1$, then $\text{Ent}_\mu(f) = H(f \cdot \mu | \mu)$ where $H$ is the relative entropy or Kullback-Leibler (KL) divergence between $f \cdot \mu$ and $\mu$.) Let $P_{SW}$ denote the transition matrix of the SW dynamics.

**Theorem 1.4.** SSM implies that there exists a constant $\delta > 0$ such that, for all cubes of $\mathbb{Z}^d$, all boundary conditions, and all functions $f : \Omega \to \mathbb{R}_+$,

$$\text{Ent}_\mu(P_{SW} f) \leq (1 - \delta) \text{Ent}_\mu(f).$$

(1.2)

It is well known (see [10] for the continuous-time analog) that, if a Markov chain satisfies relative entropy decay with rate $\delta > 0$ as in (1.2), then its mixing time is $O(\delta^{-1} \log(1/\mu_*))$, where $\mu_* = \min_\sigma \mu(\sigma)$. The previous best mixing time bound in [5] was obtained by establishing an analogous contraction for variance (instead of entropy), which yields a weaker $O(\delta^{-1} \log(1/\mu_*))$ bound on the mixing time: note that $1/\mu_* = \exp(\Theta(n))$, and so $\log(1/\mu_*) = O(n)$ whereas $\log \log(1/\mu_*) = O(\log n)$.

**Remark 1.5.** The usual approach to obtaining mixing time bounds with a $\log \log(1/\mu_*)$ dependence on $\mu_*$ is via log-Sobolev inequalities (see, e.g., [17]). The classical log-Sobolev inequality, which is equivalent to hypercontractivity, takes the form $\mathcal{D}(\sqrt{\mathcal{F}}, \sqrt{\mathcal{F}}) \geq \delta \cdot \text{Ent}_\mu(f)$, where $\mathcal{D}(\cdot, \cdot)$ is the Dirichlet form of the chain (which measures the local variation in one step). Unfortunately, however, log-Sobolev inequalities are not tight for the SW dynamics in the sense that there are functions for which $\delta$ is $\Theta(n^{-1})$; hence the best possible mixing time bound obtained in this way would be $O(n)$ (see Remark 3.2 for details). A modified log-Sobolev inequality (in continuous time) takes the form $\mathcal{D}(f, \log f) \geq \delta \cdot \text{Ent}_\mu(f)$, which is a strictly weaker (easier to satisfy) inequality, but still strong enough to establish mixing time bounds with the same dependence on $\mu_*$. Our entropy contraction bound in (1.2) actually implies the modified log-Sobolev inequality, and can be viewed as a discrete time version of it; see Section 2.

The fact that classical log-Sobolev inequalities do not capture the mixing time of the SW dynamics seems to be a more general phenomenon afflicting non-local Markov chains. These chains are popular due to their presumed speed-up over Glauber dynamics and to the fact that their updates can be parallelized. With our techniques, we are able to establish entropy contraction for another standard non-local Markov chain for the Potts model known as the alternating scan dynamics. This chain, which is used in practice to sample from the Gibbs distribution and has received some theoretical attention [5, 46, 27], also has a “bad” log-Sobolev constant, but we can show that entropy decays at a constant rate over the steps of the chain.

In one step of the alternating scan dynamics, all the even vertices (i.e., those with even coordinate sum) are updated simultaneously with a new configuration distributed according to the conditional measure on the even sub-lattice given the configuration on the odd sub-lattice; the process is then repeated for the odd vertices. The key observation is that the conditional distributions also [36] for better (sub-linear) bounds for the mixing time when $\beta \gg \beta_c(q)$. We mention that the bound in Theorem 1.3 is likely not tight, and we conjecture that the SW dynamics also mixes in $O(\log n)$ steps in this regime.
on the even and odd sub-lattices are product distributions, which makes this chain particularly amenable to parallelization and thus attractive in applications.

Let $P_E$ be the stochastic matrix corresponding to the update of the even sites conditional on the spins of the odd sites, and define $P_O$ analogously for the odd sites. The alternating scan dynamics is the Markov chain with transition matrix $S_{EO} = P_E P_O$ (or, equivalently, $S_{OE} = P_O P_E$). Note that $P_E, P_O$ do not commute (unless $\beta = 0$), so $S_{EO}$ and $S_{OE}$ are not reversible with respect to their stationary measure $\mu$. In [5] it was shown that whenever SSM holds, the mixing time of the reversibilized version $S_{OEO} = P_O P_E P_O$ of this dynamics is $O(n)$. Here we prove a much tighter bound by showing that the alternating scan dynamics itself contracts entropy at a constant rate.

**Theorem 1.6.** Let $P$ be either of the stochastic matrices $S_{EO}$ or $S_{OE}$. SSM implies that there exists a constant $\delta > 0$ such that, for all boundary conditions and all functions $f : \Omega \rightarrow \mathbb{R}_+$,

$$\text{Ent}_\mu(Pf) \leq (1 - \delta)\text{Ent}_\mu(f).$$

In particular, the Markov chain with transition matrix $P$ satisfies $T_{\text{mix}}(P) = O(\log n)$.

We note that the alternating scan dynamics is a non-local version of so-called systematic scan dynamics, a variant of Glauber dynamics in which vertices are updated in some fixed, rather than random, ordering. Due to their widespread use in practice, the effect of decay of correlations properties on the speed of convergence of this class of dynamics has been widely studied; see, e.g. [18, 28, 19]. Recently in [13], a result analogous to Theorem 1.6 was obtained for the simpler reversible dynamics with transition matrix $\frac{P_E + P_O}{2}$.

1.1 Dynamics and entropy decay in the joint space

To establish the entropy decay for the SW dynamics in Theorem 1.4, we derive a new factorization of entropy in the joint probability space on spins and edges introduced by Edwards and Sokal [21]. (This joint space is actually crucial to understanding the operation of the SW dynamics.) This new factorization has a number of interesting consequences for other natural Markov chains which we will describe shortly.

Let $\Omega_j = \Omega \times \{0, 1\}^E$ be the set of joint configurations $(\sigma, A)$ consisting of a spin assignment to the vertices $\sigma \in \Omega$ and a subset of edges $A \subseteq E$, where recall that $E$ is the set of edges with both endpoints in $V$. The Edwards-Sokal distribution on $G$ with parameters $p \in [0, 1]$ and $q \in \mathbb{N}$, and free boundary condition, is the probability measure on $\Omega_j$ given by

$$\nu(\sigma, A) = \frac{1}{Z_j} p^{|A|} (1 - p)^{|E| - |A|} \mathbf{1}(\sigma \sim A),$$

where $\sigma \sim A$ means that $A \subseteq M(\sigma)$ (i.e., that every edge in $A$ is monochromatic in $\sigma$) and $Z_j$ is the corresponding normalizing constant or partition function. When $p = 1 - e^{-\beta}$, the “spin marginal” of $\nu$ is precisely the Potts distribution $\mu$ and $Z = Z_j$; the “edge marginal” of $\nu$ corresponds to the well-known random-cluster measure; see [22, 26]. The SW dynamics alternates between spin configurations and joint spin/edge configurations.

The decay of entropy for the SW dynamics (Theorem 1.4) will be shown to be a consequence of the following theorem establishing the spin/edge factorization of entropy under SSM$^1$.

$^1$The SSM condition is the same throughout this paper, and is defined with respect to the Gibbs distribution $\mu$ for the Potts model; see Section 2 for the formal definition.
Theorem 1.7. SSM implies that there exists a constant $C \geq 1$, independent of $V$, such that for all positive functions $f : \Omega_1 \mapsto \mathbb{R}$

$$\text{Ent}_\nu(f) \leq C \left( \nu[\text{Ent}_\nu(f \mid \sigma)] + \nu[\text{Ent}_\nu(f \mid A)] \right). \quad (1.4)$$

To help understand the terms in (1.4), note that for fixed $\sigma \in \Omega$ and $A \subseteq \mathcal{E}$, $\text{Ent}_\nu(f \mid \sigma)$ and $\text{Ent}_\nu(f \mid A)$ denote the entropy of $f$ with respect to the conditional measures $\nu(\cdot \mid \sigma)$ and $\nu(\cdot \mid A)$, respectively. Therefore, $\text{Ent}_\nu(f \mid \sigma)$ and $\text{Ent}_\nu(f \mid A)$ are functions of $\sigma$ and $A$, respectively, and $\nu[\text{Ent}_\nu(f \mid \sigma)]$, $\nu[\text{Ent}_\nu(f \mid A)]$ are the corresponding expectations of the entropy functional with respect to $\nu$. We refer to the overview in Section 4 below for a more detailed interpretation of Theorem 1.7.

From the entropy factorization of the joint distribution $\nu$, we can derive Theorem 1.4 for the SW dynamics, as well as similar entropy decay inequalities for a number of natural Markov chains on the joint space for which no quantitative analysis was previously available.

First, we consider the SW dynamics in the joint space. Let $K$ denote the $|\Omega| \times |\Omega|$ stochastic matrix corresponding to re-sampling the spins of a joint configuration given the edges, and similarly let $T$ be the stochastic matrix corresponding to re-sampling the edges given the spins. The Markov chains with transition matrices $TK$, $TK$, $\frac{1}{2}(K + T)$, $KTK$ and $TKT$ are all variants of the SW dynamics in the joint space: $KT$ and $TK$ are non-reversible two-point Gibbs samplers (in the terminology of [16]), while $\frac{1}{2}(K + T)$ is their additive reversibilization and $KTK$, $TKT$ their respective multiplicative reversibilizations. We show that, under SSM, all of these dynamics satisfy entropy decay with respect to $\nu$ and hence have $O(\log n)$ mixing time.

Theorem 1.8. Let $P$ be any of the stochastic matrices $KT$, $TK$, $\frac{1}{2}(K + T)$, $KTK$ or $TKT$. SSM implies that there exists constant $\delta > 0$ such that, for all functions $f : \Omega_1 \mapsto \mathbb{R}_+$,

$$\text{Ent}_\nu(Pf) \leq (1 - \delta)\text{Ent}_\nu(f).$$

In particular, the Markov chain with transition matrix $P$ satisfies $T_{\text{mix}}(P) = O(\log n)$.

From Theorem 1.7 we can also derive tight bounds for the local (Glauber) dynamics in the joint space; this dynamics has been recently considered in [14], but as far as we know there are no results in the literature concerning its rate of convergence to stationarity. The dynamics is defined as follows: in each step, with probability $1/2$ update a vertex and with probability $1/2$ update an edge. To update a vertex, pick $v \in V$ uniformly at random and perform a “heat-bath” update at $v$ (i.e., replace the spin of $v$ with a new spin sampled from the conditional distribution of the spin at $v$ given the current spin/edge configuration); to update an edge, pick $e \in E$ uniformly at random and perform a “heat-bath” update at $e$. The local heat-bath moves, on both spins and edges, are particularly simple in the joint space; see Section 5.2. Let $P_{\text{LOCAL}}$ denote the transition matrix of this Markov chain.

Theorem 1.9. SSM implies that there exists a constant $\delta > 0$ such that, for all $f : \Omega_1 \mapsto \mathbb{R}_+$

$$\text{Ent}_\nu(P_{\text{LOCAL}}f) \leq \left(1 - \frac{\delta}{n}\right)\text{Ent}_\nu(f). \quad (1.5)$$

Moreover, the mixing time of the local dynamics satisfies $T_{\text{mix}}(P_{\text{LOCAL}}) = O(n \log n)$.

The mixing time bound in this theorem is asymptotically tight. This follows from the lower bounds in [29] by considering the projection of $P_{\text{LOCAL}}$ on the spins; see Remark 5.2.
We briefly mention several other consequences of Theorem 1.7. First, we note that Theorem 1.9 can be extended to the more general case of (weighted) block dynamics for the joint space. In addition, since the “edge marginal” of the joint measure \( \nu \) is the random-cluster distribution, we can show that the mixing time of the SW dynamics for the random-cluster model, which alternates between edge and joint configurations, is also \( O(\log n) \) for all integer \( q \geq 2 \) provided SSM holds; see Section 7 for more details about our results for random-cluster dynamics.

Finally, we note that while our results in the joint space are all stated for the free boundary condition, they actually extend to a more general class of boundary conditions that we call admissible; see Definition 4.1 in Section 4 for a definition of this class.

Organization of the paper: In Section 2 we gather definitions of some standard concepts, and various known facts used throughout the paper. Section 3 derives our algorithmic results on the Swendsen-Wang dynamics for the Potts model (Theorems 1.1 and 1.4) from our entropy factorization result (Theorem 1.7). Theorem 1.7 itself is proved in Section 4, and we go on to use it again to derive our other applications in the remaining sections. Section 5 proves entropy decay for non-local and local dynamics in the joint space (Theorems 1.8 and 1.9 respectively). Section 6 discusses the alternating scan dynamics and proves Theorem 1.6. Finally, we address the random-cluster dynamics in Section 7, concluding with a proof of Theorem 1.3.

2 Background

In this section, we formally define the spatial mixing property to be used throughout the paper. We also recall some known relations and prove some preliminary facts concerning entropy and mixing times.

2.1 Strong spatial mixing (SSM)

We assume \( V \subset \mathbb{Z}^d \) is a \( d \)-dimensional cube of \( \mathbb{Z}^d \). That is, \( V = \{0, 1, \ldots, \ell\}^d \) where \( \ell \) is a positive integer. We use \( \partial V \subset V \) to denote the internal boundary of \( V \); i.e., the set of vertices in \( V \) adjacent to at least one vertex in \( \mathbb{Z}^d \setminus V \). A boundary condition \( \psi \) for \( V \) is an assignment of spins to some (or all) vertices in \( \partial V \); i.e., \( \psi : U_\psi \rightarrow \{q\} \) with \( U_\psi \subset \partial V \). The boundary condition where \( U_\psi = \emptyset \) is called the free boundary condition. Given a boundary condition \( \psi \), each configuration \( \sigma \in \Omega \) that agrees with \( \psi \) on \( \partial V \) is assigned probability

\[
\mu_\psi(\sigma) = \frac{1}{Z_\psi} \cdot e^{-\beta |D(\sigma)|},
\]

where \( Z_\psi \) is the corresponding normalizing constant and \( D(\sigma) := \{v, w\} \in E : \sigma_v \neq \sigma_w\}. \) We define \( \mu_\psi(\sigma) = 0 \) for \( \sigma \in \Omega \) that does not agree with \( \psi \).

Let \( \mathcal{C}(V, a, b) \) be the property that, for all \( B \subset V \), all \( u \in \partial V \) and any pair of boundary conditions \( \psi, \psi_u \) on \( \partial V \) that differ only in the spin of the vertex \( u \), we have

\[
\| \mu_\psi^B - \mu_\psi^B_u \|_{TV} \leq b \exp(-a \cdot \text{dist}(u, B)),
\]  

where \( \mu_\psi^B \) and \( \mu_\psi^B_u \) are the probability measures induced in \( B \) by the Potts distribution with boundary conditions \( \psi \) and \( \psi_u \), respectively, \( \| \cdot \|_{TV} \) denotes total variation distance and \( \text{dist}(u, B) = \min_{v \in B} \|u - v\|_1 \).

Definition 2.1. We say that strong spatial mixing (SSM) holds if there exist \( a, b > 0 \) such that \( \mathcal{C}(V, a, b) \) holds for every cube \( V \subset \mathbb{Z}^d \).
We note that the definition of SSM varies in the literature, but we work here with one of the weakest (easiest to satisfy) versions. In $\mathbb{Z}^2$, this form of SSM has been established for all $q \geq 2$ and $\beta < \beta_c(q)$, where $\beta_c(q) = \ln(1 + \sqrt{q})$ is the uniqueness threshold [4, 2, 40]. Finally, we stress that the SSM property is determined only by the values of the parameters $q$ and $p = 1 - e^{-\beta}$, and not by any particular boundary condition.

**Remark 2.2.** For definiteness, we have stated all of our results for $n$-vertex $d$-dimensional cubes but they extend to more general regions of $\mathbb{Z}^d$. In particular, we can consider regions which are the union of disjoint translates of a given large enough cube. The variant of the SSM condition that requires $C(U, a, b)$ to hold for every such region $U$ is equivalent to the one in Definition 2.1 (see [38, Theorem 2.6]). As noted in [38], a version of SSM which requires $C(V, a, b)$ to hold for arbitrarily shaped regions $V$ does not hold all the way to the uniqueness threshold.

### 2.2 Mixing time, entropy, and log-Sobolev inequalities

Let $P$ be the transition matrix of an ergodic Markov chain with finite state space $\Gamma$ and stationary distribution $\pi$. Let $P^t(X_0, \cdot)$ denote the distribution of the chain after $t$ steps starting from the initial state $X_0 \in \Gamma$. The **mixing time** $T_{\text{mix}}(P)$ of the chain is defined as

$$T_{\text{mix}}(P) = \max_{X_0 \in \Gamma} \min \{ t \geq 0 : \| P^t(X_0, \cdot) - \pi \|_{\text{TV}} \leq 1/4 \}.$$ 

To prove upper bounds on the mixing time, in this paper we mostly rely on functional inequalities related to entropy.

For a function $f : \Gamma \mapsto \mathbb{R}$, let $\pi[f] = \sum_{\sigma \in \Omega} \pi(\sigma) f(\sigma)$ and $\text{Var}_\pi(f) = \pi[f^2] - \pi[f]^2$ denote its mean and variance with respect to $\pi$. Likewise, for $f$ positive, the **entropy** of $f$ with respect to $\pi$ is defined as

$$\text{Ent}_\pi(f) = \pi \left[ f \cdot \log \left( \frac{f}{\pi[f]} \right) \right] = \pi[f \cdot \log f] - \pi[f] \cdot \log \pi[f]. \quad (2.2)$$

We most often consider these functionals with respect to the Potts measure $\mu$ or the joint measure $\nu$ (as defined in (1.1) and (1.3) respectively). Sometimes, we consider the conditional expectation and the conditional entropy of functions with respect to $\nu$. In particular, if the function $f$ is such that $f : \Omega \mapsto \mathbb{R}_+$, for fixed $\sigma \in \Omega$ and $A \subseteq \mathbb{E}$, we let $\nu[f \mid \sigma] = \sum_{A \subseteq \mathbb{E}} \nu(A \mid \sigma) f(A)$, $\nu[f \mid A] = \sum_{\sigma \in \Omega} \nu(\sigma \mid A) f(\sigma, A)$ and

$$\text{Ent}_\nu(f \mid \sigma) = \nu \left[ f \cdot \log \left( \frac{f}{\nu[f \mid \sigma]} \right) \bigg| \sigma \right], \quad \text{Ent}_\nu(f \mid A) = \nu \left[ f \cdot \log \left( \frac{f}{\nu[f \mid A]} \right) \bigg| A \right].$$

Note that $\text{Ent}_\nu(f \mid \sigma)$ and $\text{Ent}_\nu(f \mid A)$ are functions of $\sigma \in \Omega$ and $A \subseteq \mathbb{E}$, respectively, and with slight abuse of notation, we write $\nu[\text{Ent}_\nu(f \mid \sigma)]$ and $\nu[\text{Ent}_\nu(f \mid A)]$ for the corresponding expectations with respect to $\nu$. The following identities hold:

$$\begin{align*}
\text{Ent}_\nu(g) &= \text{Ent}_\nu(\nu[g \mid A]) + \nu[\text{Ent}_\nu(g \mid A)]; \quad (2.3) \\
\text{Ent}_\nu(g) &= \text{Ent}_\nu(\nu[g \mid \sigma]) + \nu[\text{Ent}_\nu(g \mid \sigma)]. \quad (2.4)
\end{align*}$$

Indeed, both statements follow from the general decomposition

$$\text{Ent}_\pi(f) = \text{Ent}_\pi(\pi[f \mid \mathcal{F}]) + \pi[\text{Ent}_\pi(f \mid \mathcal{F})], \quad (2.5)$$
valid for any distribution $\pi$, and any sub $\sigma$-algebra $\mathcal{F}$, which follows by adding and subtracting the term $\pi[f \log \pi[f | \mathcal{F}]]$ in (2.2). Another basic property of entropy that we shall use is the variational principle

$$\text{Ent}_\pi(f) = \sup \{ \pi[f \varphi], \pi[e^{\varphi}] \leq 1 \},$$

valid for any distribution $\pi$, and any $f \geq 0$, where the supremum ranges over all functions $\varphi : \Gamma \to \mathbb{R}$ such that $\pi[e^{\varphi}] \leq 1$, see e.g. Proposition 2.2 in [33].

When $f \geq 0$ is such that $\pi[f] = 1$, then $\text{Ent}_\pi(f) = H(f \pi | \pi)$ corresponds to the relative entropy, or Kullback-Leibler divergence, between the distribution $f \pi$ and $\pi$.

**Definition 2.3.** A Markov chain with transition matrix $P$ and stationary distribution $\pi$ is said to satisfy the (discrete time) relative entropy decay with rate $\delta > 0$ if for all distributions $\zeta$

$$H(\zeta P | \pi) \leq (1 - \delta)H(\zeta | \pi).$$

We recall a well known consequence of entropy decay for the mixing time. For completeness, we include a proof.

**Lemma 2.4.** If a Markov chain with transition matrix $P$ and stationary distribution $\pi$ satisfies relative entropy decay with rate $\delta > 0$, then its mixing time $T_{\text{mix}}(P)$ satisfies

$$T_{\text{mix}}(P) \leq 1 + \frac{\delta^{-1}}{2} \log(8) + \log(1/\pi_*) \]$$

where $\pi_* = \min_{\sigma} \pi(\sigma)$.

**Proof of Lemma 2.4.** Pinsker’s inequality says that

$$\|\delta_{\sigma} P^n - \pi\|_2 \leq H(\delta_{\sigma} P^n | \pi),$$

where $\delta_{\sigma}(\tau) = 1(\tau = \sigma)$ is the Dirac mass at $\sigma$. Iterating (2.7),

$$\|\delta_{\sigma} P^n - \pi\|_2 \leq \frac{1}{2} (1 - \delta)^n H(\delta_{\sigma} | \pi).$$

Since $H(\delta_{\sigma} | \pi) = -\log \pi(\sigma)$ and $(1 - \delta)^n \leq e^{-\delta n}$ we obtain

$$\|\delta_{\sigma} P^n - \pi\|_2 \leq \frac{1}{4},$$

as soon as $n$ is an integer such that $n \geq \delta^{-1} \log(8 \log(1/\pi_*))$. □

**Remark 2.5.** If $\zeta$ has density $f$ with respect to $\pi$ (i.e., $\zeta = f \pi$), then $\zeta P$ has density $P^* f$ with respect to $\pi$, where $P^*$ is the adjoint or time-reversal matrix $P^*(\sigma, \sigma') = \pi(\sigma') P(\sigma, \sigma)$. Thus, (2.7) is equivalent to

$$\text{Ent}_\pi(P^* f) \leq (1 - \delta)\text{Ent}_\pi(f),$$

for all $f \geq 0$ such that $\pi[f] = 1$. By homogeneity, this is equivalent to (2.12) for all $f \geq 0$. When $P$ is reversible, that is when $P = P^*$, (2.7) is equivalent to $\text{Ent}_\pi(P f) \leq (1 - \delta)\text{Ent}_\pi(f)$ for all $f \geq 0$. 

8
The inequality (2.12) can be considered as a discrete time analogue of the so-called modified log-Sobolev inequality characterizing the relative entropy decay for continuous time Markov chains; see, e.g. [10]. Below we discuss some basic relations among (2.12), the standard log-Sobolev inequality and the modified log-Sobolev inequality.

Consider a transition matrix $P$ with stationary distribution $\pi$. The Dirichlet form associated to the pair $(P, \pi)$ is defined as
\[
D_P(f,g) = \langle f, (1 - P)g \rangle,
\] (2.13)
where $f, g$ are real functions on $\Gamma$, and $\langle f, g \rangle = \pi[fg]$ denotes the scalar product in $L^2(\pi)$. Since $f$ is real we also have
\[
D_P(f,f) = \frac{1}{2} \sum_{x,y} \pi(x)Q(x,y)(f(x) - f(y))^2,
\] (2.14)
where $Q = \frac{1}{2}(P + P^*)$. Moreover, if $P = P^*$ one has
\[
D_P(f,g) = \frac{1}{2} \sum_{x,y} \pi(x)P(x,y)(f(x) - f(y))(g(x) - g(y)),
\] (2.15)
for all $f, g$.

**Definition 2.6.** The pair $(P, \pi)$ is said to satisfy the (standard) log-Sobolev inequality (LSI) with constant $\alpha$ if for all $f \geq 0$:
\[
D_P(\sqrt{f},\sqrt{f}) \geq \alpha \text{Ent}_\pi f.
\] (2.16)

It is said to satisfy the modified log-Sobolev inequality (MLSI) with constant $\delta$ if for all $f \geq 0$:
\[
D_P(f, \log f) \geq \delta \text{Ent}_\pi f.
\] (2.17)

It is well known that the Log-Sobolev inequality is equivalent to the so-called hypercontractivity (see [17, Theorem 3.5]), while the modified Log-Sobolev inequality (2.17) is equivalent to exponential decay of the relative entropy with rate $\delta$ for the continuous time kernel $K_t = e^{(P - 1)t}$ (see [17, Theorem 3.6]). Note that we are not assuming reversibility. To see the relation between the MLSI and the entropy decay in continuous time, note that if $K_t = e^{(P - 1)t}$ and $f$ has mean $\pi[f] = 1$ then using $K_t^* = e^{(P^* - 1)t}$ one checks that the time derivative of the relative entropy satisfies
\[
\frac{d}{dt} H(\zeta K_t | \pi) = \frac{d}{dt} \text{Ent}(K_t^* f) = -D_P(K_t^* f, \log K_t^* f),
\] (2.18)
where $\zeta = f \cdot \pi$. Therefore (2.17) implies, for all $t \geq 0$:
\[
H(\zeta K_t | \pi) \leq H(\zeta | \pi)e^{-\delta t}.
\]

Next, we observe that the bound (2.12) is stronger than the MLSI in (2.17).

**Lemma 2.7.** If the entropy decay holds with rate $\delta$ in discrete time then it holds with the same rate in continuous time. That is, (2.12) implies the MLSI with constant $\delta$. 

9
Therefore, \[ \alpha \text{satisfies the LSI with constant } \delta \text{ time with rate } \mu. \]

Proof. The first assertion is proved in [\ref{17}, Lemma 2.7]. Here we recall a result of Miclo [\ref{43}] showing in what sense the LSI implies the discrete time entropy decay.

Lemma 2.8. If the pair \((P^*, \pi)\) satisfies the standard LSI with constant \(\alpha\), then the discrete time entropy decay holds for \((P, \pi)\) with constant \(\delta = \alpha\). In particular, if \(P\) is reversible and \((P, \pi)\) satisfies the LSI with constant \(\alpha\), then for all \(f \geq 0\):

\[
\text{Ent}_\pi Pf \leq (1 - \alpha)\text{Ent}_\pi f. \tag{2.19}
\]

Proof. The first assertion is proved in [\ref{43}, Proposition 6]. The second assertion follows from the first and the simple observation that if \(P = P^*\) then the LSI for \((P, \pi)\) implies the LSI for \((P^* P, \pi)\) with the same constant since \(P^* P = P^2 \leq P\) as quadratic forms in \(L^2(\pi)\).}

3 Entropy decay of the SW dynamics

In this section we show how to obtain our main algorithmic result on the mixing time of the SW dynamics (Theorem 1.1) from our decay of entropy result (Theorem 1.7). We start by proving Theorem 1.4 from which Theorem 1.1 follows.

Proof of Theorem 1.4. We are required to show that for all positive functions \(f : \Omega \mapsto \mathbb{R}_+\) with \(\mu[f] = 1\), one has

\[
\text{Ent}_\mu(P_{sw} f) \leq (1 - \delta)\text{Ent}_\mu(f). \tag{3.1}
\]

The transition matrix of the SW dynamics satisfies \(P_{sw}(\sigma, \tau) = \sum_{A \subseteq M(\sigma)} \nu(A \mid \sigma) \nu(\tau \mid A)\), where we recall that \(M(\sigma)\) is the set of monochromatic edges in \(\sigma\). Hence,

\[
P_{sw} f(\sigma) = \sum_{\tau \in \Omega} P_{sw}(\sigma, \tau) f(\tau) = \sum_{\tau \in \Omega} \sum_{A \subseteq M(\sigma)} \nu(A \mid \sigma) \nu(\tau \mid A) f(\tau)
\]

\[
= \sum_{\tau \in \Omega} \sum_{A \subseteq M(\sigma)} \nu(A \mid \sigma) \nu(\tau \mid A) \hat{f}(\tau, A),
\]

where the function \(\hat{f} : \Omega \mapsto \mathbb{R}_+\) is the “lift” of \(f\) to the joint space, i.e., \(\hat{f}(\sigma, A) = f(\sigma)\) for every \((\sigma, A) \in \Omega_j\). Recalling that we write \(\nu[f], \nu[f|A], \nu[f|\sigma]\) for the expectations of \(f\) with respect to the measures \(\nu(\cdot), \nu(\cdot \mid A), \nu(\cdot \mid \sigma)\), respectively, we obtain

\[
P_{sw} f(\sigma) = \sum_{A \subseteq M(\sigma)} \nu(A \mid \sigma) \nu[\hat{f} \mid A] = \nu[\nu[\hat{f} \mid A] \mid \sigma] = \nu[g \mid \sigma],
\]
where for ease of notation we set \( g := \nu[f | A] \). Since \( \mu[f] = 1 \), we have \( \mu[P_{sw}f] = 1 \) and
\[
\text{Ent}_\mu(P_{sw}f) = \mu[(P_{sw}f) \log(P_{sw}f)] = \mu[\nu[g | \sigma] \log(\nu[g | \sigma])].
\]
The convexity of the function \( x \cdot \log x \) and Jensen’s inequality imply
\[
\nu[g | \sigma] \log(\nu[g | \sigma]) \leq \nu[g \log g | \sigma],
\]
and then
\[
\text{Ent}_\mu(P_{sw}f) \leq \mu[\nu[g \log g | \sigma]] = \nu[\nu[g \log g | \sigma] = \nu[g \log g] = \text{Ent}_\nu(g),
\]
since \( \nu[g] = \nu[\hat{f}] = \mu[f] = 1 \).

For any function \( h : \Omega_i \mapsto \mathbb{R}_+ \), we have by (2.3) \( \text{Ent}_\nu(h) = \text{Ent}_\nu(\nu[h | A]) + \nu[\text{Ent}_\nu(h | A)] \). Hence,
\[
\text{Ent}_\nu(\hat{f}) = \text{Ent}_\nu(g) + \nu[\text{Ent}_\nu(\hat{f} | A)],
\]
which by (3.2) gives
\[
\text{Ent}_\mu(P_{sw}f) \leq \text{Ent}_\nu(\hat{f}) - \nu[\text{Ent}_\nu(\hat{f} | A)].
\]
The function \( \hat{f} \) depends on \( \sigma \) only, so \( \text{Ent}_\nu(\hat{f} | \sigma) = 0 \). Therefore,
\[
\text{Ent}_\mu(P_{sw}f) \leq \text{Ent}_\nu(\hat{f}) - \nu\left[\text{Ent}_\nu(\hat{f} | A) + \text{Ent}_\nu(\hat{f} | \sigma)\right].
\]
We may now apply Theorem 1.7 to obtain
\[
\text{Ent}_\mu(P_{sw}f) \leq (1 - \delta)\text{Ent}_\nu(\hat{f}),
\]
for some constant \( \delta \in (0, 1] \). Inequality (3.1) follows from the fact that \( \text{Ent}_\nu(\hat{f}) = \text{Ent}_\mu(f) \). \qed

We can now provide the proof of Theorem 1.1.

Proof of Theorem 1.1. The SW dynamics is reversible with respect to \( \mu \) and so \( P_{sw} = P_{sw}^* \). Then, it follows from Theorem 1.4, Lemma 2.4 and Remark 2.5 that the mixing time of the SW dynamics is \( O(\log n) \) whenever SSM holds. \qed

Remark 3.1. The previous proof applies to the Potts measure \( \mu \) obtained as the marginal on spins of the joint measure \( \nu \). If \( \nu \) is as in (1.3) this yields only the Potts measure on \( V \) with the free boundary condition. However, using the more general result in Theorem 4.2 below, we can obtain the Potts measure with any boundary condition by choosing a pure spin boundary condition for \( \nu \); see also Definition 4.1 and the examples immediately following it. Therefore, Theorem 1.1 holds for arbitrary boundary conditions, as stated.

Remark 3.2. As mentioned in Section 2, the entropy contraction bound in (1.2) implies a modified log-Sobolev inequality, and can be viewed as a discrete time version of it. The classical log-Sobolev constant, however, is not tight for the SW dynamics. Indeed, the remark in [37, Section 3.7] shows a test function \( f \) such that \( \text{Var}_\mu(\sqrt{f})/\text{Ent}_\mu(f) = O(n^{-1}) \). Since \( D_{P_{sw}}(\sqrt{f}, \sqrt{f}) = \nu[\text{Var}(\sqrt{f} | A)] \), it follows from monotonicity of variance that \( D_{P_{sw}}(\sqrt{f}, \sqrt{f}) \leq \text{Var}_\mu(\sqrt{f}) \) and so
\[
\frac{D_{P_{sw}}(\sqrt{f}, \sqrt{f})}{\text{Ent}_\mu(f)} = O(n^{-1}).
\]
4 Factorization of entropy in the joint space

In this section we prove our main technical result, Theorem 1.7, which expresses the factorization of entropy in the joint spin/edge space. For simplicity, we have so far stated our results for the joint space only for the free boundary condition, corresponding to the measure \( \nu \) in (1.3). However, all of our results apply to a more general class of boundary conditions we call admissible.

**Admissible boundary conditions.** Let \( \partial V \) be the set of vertices of \( V \) with a neighbor in \( \mathbb{Z}^d \setminus V \). Let \( \partial E \) denote the set of edges in \( E \) with at least one endpoint in \( \partial V \). (Recall that \( E \) is the set of edges with both endpoints in \( V \).) We consider boundary conditions for the joint space on subsets \( V_0 \subseteq \partial V \) and \( E_0 \subseteq \partial E \). Specifically, we let \( \psi : V_0 \mapsto [q] \) and \( \varphi : E_0 \mapsto \{0,1\} \) and define

\[
\nu^{\psi,\varphi}(\sigma, A) = \frac{1}{Z^{\psi,\varphi}} p^{|A|} (1 - p)^{|E| - |A|} \mathbf{1}(\sigma \sim A) \mathbf{1}(A \sim \varphi),
\]

where \( \sigma \sim A \) means that \( A \subseteq M(\sigma) \), \( \sigma \sim \psi \) that \( \sigma \) and \( \psi \) agree on the spins in \( V_0 \), and \( A \sim \varphi \) that \( A \) and \( \varphi \) agree on the edges in \( E_0 \). As usual, \( Z^{\psi,\varphi} \) is the corresponding normalizing constant, or partition function.

**Definition 4.1.** We call the boundary condition admissible if \( E_0 \subseteq \{\{u,v\} \in \partial E : u \in V_0\} \); that is, if all edges in \( E_0 \) have at least one endpoint in \( V_0 \).

Notice that the free boundary condition \((V_0 = \emptyset \text{ and } E_0 = \emptyset)\) is admissible, and all pure spin boundary conditions \((V_0 \subset \partial V \text{ and } E_0 = \emptyset)\) are also admissible. In this case, the marginal on spins is just the Potts measure with \( \psi \) as the boundary condition on \( \partial V \) with \( U^{\psi} = V_0 \).

The main motivation for introducing the notion of admissible boundary conditions is that it guarantees that the spin marginal of \( \nu^{\psi,\varphi} \) has the desired exponential decay of correlations if the parameters \( q \) and \( \beta \) are such that SSM holds. In fact, we shall see that SSM (on the spin marginal) is the only requirement for the edge/spin entropy factorization to hold in the joint space. Consequently, all of our results concerning the joint measure and its dynamics extend to the more general class of admissible boundary conditions. We can therefore restate Theorem 1.7 from the introduction as follows.

**Theorem 4.2.** Let \( \nu := \nu^{\psi,\varphi} \) be the joint distribution with an admissible boundary condition \((\psi, \varphi)\). If \( q \) and \( \beta \) are such that SSM holds, there exists a constant \( C \geq 1 \) independent of \( V \) such that, for all positive functions \( f : \Omega_1 \mapsto \mathbb{R}_+ \),

\[
\text{Ent}_\nu(f) \leq C \left( \nu[\text{Ent}_\nu(f \mid \sigma)] + \nu[\text{Ent}_\nu(f \mid A)] \right).
\]

For simplicity, we shall continue to write \( \nu \) for the joint measure \( \nu^{\psi,\varphi} \) and \( \mu \) for its marginal on spins. In fact, we shall see that our proofs in this section are largely oblivious to the boundary condition. We also remark that, while we could allow a slightly more general family of boundary conditions than the admissible ones, some limitations are needed. For instance, arbitrary edge boundary conditions are known to cause long-range dependencies; see, e.g., [9, 7]. We proceed next with the proof of Theorem 4.2.

4.1 Proof of Theorem 4.2

Let \( E \) denote the set of even vertices in \( V \) and \( O \) the set of odd vertices in \( V \). Let \( \nu(\cdot \mid \sigma_E, A) \) denote the measure \( \nu \) conditioned on \( \sigma_E = \{\sigma_v, v \in E\} \) and \( A \subseteq E \). Similarly, \( \nu(\cdot \mid \sigma_O, A) \) denotes the measure \( \nu \) conditioned on \( \sigma_O = \{\sigma_v, v \in O\} \) and \( A \). We use \( \text{Ent}_\nu(f \mid \sigma_E, A) \) and \( \text{Ent}_\nu(f \mid \sigma_O, A) \) to
denote the corresponding conditional entropies and \( \nu [\text{Ent}_{\nu}(f \mid \sigma_E, A)] \), \( \nu [\text{Ent}_{\nu}(f \mid \sigma_O, A)] \) for their expectations with respect to \( \nu \).

**Overview.** The following high level observations might be of help before entering the technical details of the proof. First, notice that the statement in the theorem would trivially hold true with constant \( C = 1 \) if \( \nu \) were a product measure with respect to the two sets of variables \( (\sigma, A) \). This is a consequence of standard factorization properties of product measures. Thus, the minimal constant \( C \) for which that statement holds is a measure of the “cost” for “separating” the two sets of variables. When the dependencies between the two sets of variables are very weak, a factorization statement could be obtained as in [15]. However, in our case the dependencies are not weak, since the spin variables interact locally with the edge variable in a strong way. For instance, the presence of the edge \( xy \) in \( A \) forces deterministically the condition \( \sigma_x = \sigma_y \). Thus, the fact that our statement holds with a constant \( C \) independent of \( n \) is highly nontrivial. On the other hand, for every \( x \in V \) one can separate locally the two variables \((\sigma_x, A_x)\), where \( A_x \) denotes the set of edge variables for edges incident to \( x \), by paying a finite cost \( C \); this is the content of Lemma 4.8 below. By adapting a technique introduced recently in [13] we can lift this local factorization to a global factorization statement for the conditional measure \( \nu(\cdot \mid \sigma_E) \), respectively \( \nu(\cdot \mid \sigma_O) \), obtained by conditioning on the spin variables of all even vertices \( E \subset V \), respectively of all odd vertices \( O \subset V \). This is the content of Lemma 4.5. It relies crucially on the fact that \( \nu(\cdot \mid \sigma_E) \) is a product measure with respect to \( \{ (\sigma_x, A_x), x \in O \} \), and \( \nu(\cdot \mid \sigma_O) \) is a product measure with respect to \( \{ (\sigma_x, A_x), x \in E \} \). Thus, we reduce the problem of separating the spin/edge variables \((\sigma, A)\) to the problem of separating the even/odd spin variables \((\sigma_E, \sigma_O)\). We then conclude by adapting to our setting one of the main results of [13] which allows precisely this even/odd factorization. This is the content of Lemma 4.6.

We now turn to the actual proof. We start by stating a key factorization of entropy into even and odd sites that was proved for \( \mu \) in [13].

**Theorem 4.3 (Theorem 4.3 in [13]).** SSM implies that there exists a constant \( \delta > 0 \) such that for all cubes of \( \mathbb{Z}^d \), all boundary conditions, and for all functions \( f : \Omega \mapsto \mathbb{R}_+ \),

\[
\mu [\text{Ent}_{\mu}(f \mid \sigma_E) + \text{Ent}_{\mu}(f \mid \sigma_O)] \geq \delta \text{Ent}_{\mu}(f).
\]

The next simple lemma shows that conditioning on the spin configuration of the even or the odd sub-lattice can only decrease the entropy of a function with respect to \( \nu(\cdot \mid A) \).

**Lemma 4.4.** For all positive functions \( f : \Omega \mapsto \mathbb{R}_+ \) we have

\[
\nu [\text{Ent}_{\nu}(f \mid A)] \geq \nu [\text{Ent}_{\nu}(f \mid \sigma_E, A)]; \quad \text{and}
\]

\[
\nu [\text{Ent}_{\nu}(f \mid A)] \geq \nu [\text{Ent}_{\nu}(f \mid \sigma_O, A)].
\]

**Proof.** From the decomposition (2.5) applied to \( \pi = \nu(\cdot \mid A) \) with \( \mathcal{F} \) the \( \sigma \)-algebra generated by \( \sigma_E \), we can write

\[
\text{Ent}_{\nu}(f \mid A) = \text{Ent}_{\nu}(f \mid \sigma_E, A) + \nu [f \mid \sigma_E, A] \mid A \geq \text{Ent}_{\nu}(f \mid \sigma_E, A).
\]

Taking the expectation with respect to \( \nu \) we obtain

\[
\nu [\text{Ent}_{\nu}(f \mid A)] \geq \nu [\text{Ent}_{\nu}(f \mid \sigma_E, A)].
\]

The same argument applies to the odd sites to deduce that \( \nu [\text{Ent}_{\nu}(f \mid A)] \geq \nu [\text{Ent}_{\nu}(f \mid \sigma_O, A)] \). \( \square \)
The advantage of working with \( \nu(\cdot | \sigma_O, A) \) or \( \nu(\cdot | \sigma_E, A) \) instead of \( \nu(\cdot | A) \) is that once we condition on the spins on all odd (resp. even) sites the measure becomes a product over the even (resp. odd) vertices, and we can exploit tensorization properties of entropy for product measures. The next lemma is a key step in the proof.

**Lemma 4.5.** There exists a constant \( \delta_1 > 0 \) such that, for all functions \( f : \Omega_j \mapsto \mathbb{R}_+ \),

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma) \right] + \nu \left[ \text{Ent}_\nu(f \mid \sigma_O, A) \right] \geq \delta_1 \nu \left[ \text{Ent}_\nu(f \mid \sigma_O) \right], \tag{4.2}
\]

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma) \right] + \nu \left[ \text{Ent}_\nu(f \mid \sigma_E, A) \right] \geq \delta_1 \nu \left[ \text{Ent}_\nu(f \mid \sigma_E) \right]. \tag{4.3}
\]

We defer the proof of Lemma 4.5 to later. Adding up (4.2) and (4.3) and using Lemma 4.5 we obtain the estimate

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma) + \text{Ent}_\nu(f \mid A) \right] \geq \frac{\delta_1}{2} \nu \left[ \text{Ent}_\nu(f \mid A) \right]. \tag{4.4}
\]

We then use a generalization of the entropy factorization in Theorem 4.3 to reconstruct, in the presence of SSM, the global entropy \( \text{Ent}_\nu(f) \) from the conditional entropies \( \nu \left[ \text{Ent}_\nu(f \mid \sigma_E) \right] \) and \( \nu \left[ \text{Ent}_\nu(f \mid \sigma_O) \right] \) on the right hand side of (4.4).

**Lemma 4.6.** SSM implies that there exists a constant \( \delta_2 > 0 \) such that, for all positive functions \( f : \Omega_j \mapsto \mathbb{R}_+ \),

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma_E) + \text{Ent}_\nu(f \mid \sigma_O) \right] \geq \delta_2 \nu \left[ f \right]. \tag{4.5}
\]

**Proof.** If \( f \) depends only on the spins \( \sigma \) then the result coincides with the factorization into even and odd sites in Theorem 4.3. However, here \( f \) depends on the edge configuration \( A \) as well. We need the following observations:

\[
\text{Ent}_\nu(f \mid \sigma_O) = \text{Ent}_\nu(\nu[f \mid \sigma] \mid \sigma_O) + \nu \left[ \text{Ent}_\nu(f \mid \sigma) \mid \sigma_O \right], \tag{4.5}
\]

\[
\text{Ent}_\nu(f \mid \sigma_E) = \text{Ent}_\nu(\nu[f \mid \sigma] \mid \sigma_E) + \nu \left[ \text{Ent}_\nu(f \mid \sigma) \mid \sigma_E \right]. \tag{4.6}
\]

Indeed, (4.5) and (4.6) follow from the decomposition (2.5) and the fact that \( \nu[\cdot : \mid \sigma_E, \sigma_O] = \nu[\cdot : \mid \sigma] \).

Now, since the function \( \nu[f \mid \sigma] \) depends only on the spin configuration \( \sigma \),

\[
\nu \left[ \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_E) + \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_O) \right] = \mu \left[ \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_E) + \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_O) \right],
\]

and we may apply Theorem 4.3 to the function \( \nu[f \mid \sigma] \). This theorem says that if SSM holds, then there exists a constant \( \delta_2 \in (0, 1] \) such that

\[
\mu \left[ \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_E) + \text{Ent}_\mu(\nu[f \mid \sigma] \mid \sigma_O) \right] \geq \delta_2 \text{Ent}_\mu(\nu[f \mid \sigma]). \tag{4.7}
\]

Therefore, observing that

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma) \mid \sigma_O \right] + \nu \left[ \text{Ent}_\nu(f \mid \sigma) \mid \sigma_E \right] = 2 \nu \left[ \text{Ent}_\nu(f \mid \sigma) \right],
\]

we obtain from (4.5), (4.6) and (4.7)

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma_E) + \text{Ent}_\nu(f \mid \sigma_O) \right] \geq \delta_2 \text{Ent}_\nu(\nu[f \mid \sigma]) + 2 \nu \left[ \text{Ent}_\nu(f \mid \sigma) \right].
\]

Since \( \delta_2 \leq 1 \), the standard decomposition in (2.4) implies

\[
\nu \left[ \text{Ent}_\nu(f \mid \sigma_E) + \text{Ent}_\nu(f \mid \sigma_O) \right] \geq \delta_2 \text{Ent}_\nu(f),
\]

as claimed. \( \square \)
The proof of Theorem 4.2 is now immediate.

Proof of Theorem 4.2. Apply inequality (4.4) and Lemma 4.6 with \( C = 2/\delta_1 \delta_2 \).

\[ \square \]

**Remark 4.7.** We remark that the only part of the proof where the SSM assumption is used is Lemma 4.6. This is also the only place where we use the geometry of \( \mathbb{Z}^d \). The rest of the analysis, namely the inequality (4.4), is valid for any bipartite graph \( G \), for any \( \beta \) and any integer \( q \geq 2 \).

It remains for us to provide the proof of Lemma 4.5, which we do in the next subsection.

### 4.2 Proof of Lemma 4.5

Before giving the proof of Lemma 4.5, we mention several useful facts about the joint distribution \( \nu \). The first key fact is that, for any fixed configuration \( \sigma_O \) of spins on the odd sub-lattice, the conditional probability \( \nu(\cdot \mid \sigma_O) \) is a product measure. That is,

\[ \nu(\cdot \mid \sigma_O) = \bigotimes_{x \in E} \nu_x(\cdot \mid \sigma_O), \tag{4.8} \]

where, for each \( x \in E \), \( \nu_x(\cdot \mid \sigma_O) \) is the probability measure on \( \{1, \ldots, q\} \times \{0,1\}^{\deg(x)} \), where \( \deg(x) \) denotes the degree of \( x \), described as follows: pick the spin of site \( x \) according to the Potts measure on \( x \) conditioned on the spin of its neighbors in \( \sigma_O \); then, independently for every edge \( xy \in E \) incident to the vertex \( x \), if \( \sigma_x = \sigma_y \) set \( A_{xy} = 1 \) with probability \( p \) and set \( A_{xy} = 0 \) otherwise; if \( \sigma_x \neq \sigma_y \), set \( A_{xy} = 0 \). (Note that in this section, to simplify notation, we shall use \( x \) to denote the edge \( \{x,y\} \), and view the edge configuration \( A \) as a vector in \( \{0,1\}^E \).

Consider now the measure \( \nu(\cdot \mid \sigma_O, A) \) obtained by further conditioning on a valid configuration of all edge variables \( A \). Here \( A \) is valid if it is compatible with the fixed spins \( \sigma_O \). This is again a product measure; namely

\[ \nu(\cdot \mid \sigma_O, A) = \bigotimes_{x \in E} \nu_x(\cdot \mid \sigma_O, A), \tag{4.9} \]

where \( \nu_x(\cdot \mid \sigma_O, A) \) is the probability measure on \( \{1, \ldots, q\} \) that is uniform if \( x \) has no incident edges in \( A \), and is concentrated on the unique admissible value given \( \sigma_O \) and \( A \) otherwise.

Next, we note that \( \nu(\cdot \mid \sigma) \) is a product of Bernoulli\((p)\) random variables over all monochromatic edges in \( \sigma \), while it is concentrated on \( A_e = 0 \) on all remaining edges. Therefore we may write

\[ \nu(\cdot \mid \sigma) = \bigotimes_{x \in E} \nu_x(\cdot \mid \sigma), \tag{4.10} \]

where \( \nu_x(\cdot \mid \sigma) \) is the probability measure on \( \{0,1\}^{\deg(x)} \) given by the product of Bernoulli\((p)\) variables on all edges \( xy \) incident to \( x \) such that \( \sigma_x = \sigma_y \) and is concentrated on \( A_{xy} = 0 \) if \( \sigma_x \neq \sigma_y \).

We write \( \text{Ent}_x(\cdot \mid \sigma_O), \text{Ent}_x(\cdot \mid \sigma_O, A), \text{Ent}_x(\cdot \mid \sigma) \) for the entropies with respect to the distributions \( \nu_x(\cdot \mid \sigma_O), \nu_x(\cdot \mid \sigma_O, A), \nu_x(\cdot \mid \sigma) \) respectively. The first observation is that, for every site \( x \), there is a local factorization of entropies in the following sense.

**Lemma 4.8.** There exists a constant \( \delta_1 > 0 \) such that, for all functions \( f \geq 0 \) and all \( x \in E \),

\[ \nu_x[\text{Ent}_x(f \mid \sigma) \mid \sigma_O] + \nu_x[\text{Ent}_x(f \mid \sigma_O, A) \mid \sigma_O] \geq \delta_1 \text{Ent}_x(f \mid \sigma_O). \tag{4.11} \]
**Proof.** For \( x \in V \), let \( A_x \) be random variable in \( \{0,1\}^{\deg(x)} \) corresponding to the configuration of the edges incident to \( x \) in \( A \). If we replace entropy by variance, then (4.11) is a spectral gap inequality for the Markov chain where the variable \((\sigma_x, A_x) \in [q] \times \{0,1\}^{\deg(x)} =: S \) is updated as follows. At each step, with probability \( 1/2 \) the spin \( \sigma_x \) is updated with a sample from \( \nu_x(\cdot | \sigma_x, A_x) \), and with probability \( 1/2 \) the edges \( A_x \) incident to \( x \) are simultaneously updated with a sample from \( \nu_x(\cdot | \sigma) \). Let \( P_x = \frac{Q_x + S_x}{2} \) denote the transition matrix of this Markov chain, where \( Q_x \), \( S_x \) are the stochastic matrices corresponding to the spin and edge moves at \( x \), respectively. Let \( D_{P_x} = \frac{1}{2} D_{Q_x} + \frac{1}{2} D_{S_x} \) denote the corresponding Dirichlet form. Observe that, by updating first the edges with an empty configuration and then the spin, two arbitrary initial configurations can be coupled after two steps with probability at least \( \frac{1}{4}(1-p)^{-2d} \), and thus an immediate coupling argument shows that the spectral gap of the transition matrix \( P_x \) is at least \( \delta_0 \) where \( \delta_0 > 0 \) is a constant depending only on \( p \) and \( d \). Therefore, for any function \( f : S \mapsto \mathbb{R}_+ \)

\[
D_{P_x}(f, f) \geq \delta_0 \text{Var}_x(f | \sigma_O).
\]

Noticing that \( D_{Q_x}(f, f) = \nu_x[\text{Var}_x(f | \sigma) | \sigma_O] \) and \( D_{S_x}(f, f) = \nu_x[\text{Var}_x(f | \sigma_O, A) | \sigma_O] \), we arrive at the inequality

\[
\nu_x[\text{Var}_x(f | \sigma) | \sigma_O] + \nu_x[\text{Var}_x(f | \sigma_O, A) | \sigma_O] \geq 2\delta_0 \text{Var}_x(f | \sigma_O).
\]

(4.12)

A well known general relation between entropy and variance (see, e.g., Theorem A.1 and Corollary A.4 in [17]) shows that, for all \( f \geq 0 \),

\[
\text{Ent}_x(f | \sigma_O) \leq C_1 \text{Var}_x(\sqrt{f} | \sigma_O),
\]

(4.13)

where \( C_1 \) is a constant depending only on \( q, p \) and \( d \). Thus, applying (4.12) to \( \sqrt{f} \) instead of \( f \), we obtain

\[
\nu_x[\text{Var}_x(\sqrt{f} | \sigma) | \sigma_O] + \nu_x[\text{Var}_x(\sqrt{f} | \sigma_O, A) | \sigma_O] \geq \frac{2\delta_0}{C_1} \text{Ent}_x(f | \sigma_O).
\]

(4.14)

The conclusion (4.11) follows by recalling that for any \( f \geq 0 \) the variance of \( \sqrt{f} \) is at most the entropy of \( f \) for any underlying probability measure; see, e.g., [32, Lemma 1]. In particular, \( \text{Var}_x(\sqrt{f} | \sigma) \leq \text{Ent}_x(f | \sigma) \) and \( \text{Var}_x(\sqrt{f} | \sigma_O, A) \leq \text{Ent}_x(f | \sigma_O, A) \). \( \square \)

To prove Lemma 4.5, we need to lift the inequality of Lemma 4.8 to the product measure \( \nu(\cdot | \sigma_O) = \otimes_{x \in E} \nu_x(\cdot | \sigma_O) \).

**Proof of Lemma 4.5.** We will prove (4.2); exactly the same argument applies to (4.3). Let \( x = 1, \ldots, n \) denote an arbitrary ordering of the even sites \( x \in E \). Let \( A_x \in \{0,1\}^{\deg(x)} \) be the random variable corresponding to the state of the edges incident to \( x \). We write \( \xi_x = (\sigma_x, A_x) \) for the pair of variables at \( x \). We first observe that

\[
\text{Ent}_x(f | \sigma_O) = \sum_{x=1}^{n} \nu[\text{Ent}_x(g_{x-1} | \sigma_O) | \sigma_O],
\]

(4.15)

where \( g_x = \nu[f | \sigma_O, \xi_{x+1}, \ldots, \xi_n] \), so that \( g_0 = f \) and \( g_n = \nu[f | \sigma_O] \). To prove (4.15), we note that since \( \nu(\cdot | \sigma_O) = \otimes_{x \in E} \nu_x(\cdot | \sigma_O) \), one has \( \nu_x[g_{x-1} | \sigma_O] = g_x \). Therefore,

\[
\text{Ent}_x(f | \sigma_O) = \nu [g_0 \log (g_0/g_n) | \sigma_O] = \sum_{x=1}^{n} \nu [g_0 \log (g_{x-1}/g_x) | \sigma_O].
\]
Since the \( g_x \) are (conditional) expectations, we deduce

\[
\Ent_{\nu}(f \mid \sigma_O) = \sum_{x=1}^{n} \nu \left[ g_{x-1} \log \left( \frac{g_{x-1}}{g_x} \right) \mid \sigma_O \right]
\]

\[
= \sum_{x=1}^{n} \nu \left[ \nu_x \left[ g_{x-1} \log \left( \frac{g_{x-1}}{g_x} \right) \mid \sigma_O \right] \right] \mid \sigma_O
\]

\[
= \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma_O) \right] \mid \sigma_O].
\] (4.16)

From (4.16), using Lemma 4.8 we obtain

\[
\delta_1 \Ent_{\nu}(f \mid \sigma_O) \leq \sum_{x=1}^{n} \nu \left[ \nu_x \left[ \Ent_x(g_{x-1} \mid \sigma) \right] \mid \sigma_O \right] + \nu_x \left[ \Ent_x(g_{x-1} \mid \sigma_O, A) \right] \mid \sigma_O]
\]

\[
= \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma) \right] + \Ent_x(g_{x-1} \mid \sigma_O, A) \mid \sigma_O].
\] (4.17)

Observe that \( \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma) \right] \mid \sigma_O \) and \( \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma_O, A) \right] \mid \sigma_O \) are “tensorized” versions of \( \nu \left[ \Ent_{\nu}(f \mid \sigma) \right] \mid \sigma_O \) and \( \nu \left[ \Ent_{\nu}(f \mid \sigma_O, A) \right] \), respectively, which are the terms on the right hand side of (4.2). Using similar but somewhat more involved ideas to those used to derive (4.16), we can establish the following.

**Lemma 4.9.**

1. \( \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma) \right] \mid \sigma_O \leq \nu \left[ \Ent_{\nu}(f \mid \sigma) \right] \mid \sigma_O \)

2. \( \sum_{x=1}^{n} \nu \left[ \Ent_x(g_{x-1} \mid \sigma, A) \right] \mid \sigma_O \leq \nu \left[ \Ent_{\nu}(f \mid \sigma_O, A) \right] \mid \sigma_O \)

Before providing the proof of this lemma, we finish the proof of Lemma 4.5. Inequality (4.17) together with parts 1 and 2 of Lemma 4.9 show that

\[
\delta_1 \Ent_{\nu}(f \mid \sigma_O) \leq \nu \left[ \Ent_{\nu}(f \mid \sigma) \right] \mid \sigma_O \right] + \nu \left[ \Ent_{\nu}(f \mid \sigma_O, A) \right] \mid \sigma_O].
\] (4.18)

Taking expectations with respect to \( \nu \) in (4.18) we arrive at (4.2) and the proof is complete. \( \square \)

We finish the proof of Lemma 4.5 by providing the proof of Lemma 4.9.

**Proof of Lemma 4.9.** We start with part 2. Let \( h_x = \nu \left[ f \mid \sigma_O, \sigma_{x+1}, \ldots, \sigma_n, A \right] \), so that \( h_0 = f \) and \( h_n = \nu \left[ f \mid \sigma_O, A \right] \). Since \( \nu(\cdot \mid \sigma_O, A) \) is a product measure, \( \nu_x[h_{x-1} \mid \sigma_O, A] = h_x \). Therefore, reasoning as in (4.15) we obtain

\[
\Ent_{\nu}(f \mid \sigma_O, A) = \sum_{x=1}^{n} \nu \left[ \Ent_x(h_{x-1} \mid \sigma_O, A) \right] \mid \sigma_O, A] \right].
\] (4.19)

Taking expectations with respect to \( \nu(\cdot \mid \sigma_O) \) in (4.19) we see that it is sufficient to show that, for all \( x \),

\[
\nu \left[ \Ent_x(g_{x-1} \mid \sigma_O, A) \right] \mid \sigma_O] \leq \nu \left[ \Ent_x(h_{x-1} \mid \sigma_O, A) \right] \mid \sigma_O].
\] (4.20)
To prove (4.20), we introduce the measures \( \mu_k = \otimes_{x=1}^k \nu_x(\cdot | \sigma_O) \) and \( \mu_k^A = \otimes_{x=1}^k \nu_x(\cdot | \sigma_O, A) \). Then we have \( g_x = \mu_x[f], h_x = \mu_x^A[f], \) and \( g_x = \mu_x[h_x] \). Also, we simplify the notation by writing \( \nu_x(\cdot | \sigma_O, A) =: \nu_x^A \). Now the product structure implies the commutation relation

\[
\nu_x^A g_{x-1} = \nu_x^A \mu_{x-1} h_{x-1} = \mu_{x-1} \nu_x^A h_{x-1}. \tag{4.21}
\]

Therefore,

\[
\nu \left[ \operatorname{Ent}_x(g_{x-1} | \sigma_O, A) | \sigma_O \right] = \nu \left[ g_{x-1} \log \left( g_{x-1}/\nu_x^A g_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ \mu_{x-1} h_{x-1} \log \left( \mu_{x-1} h_{x-1}/\mu_{x-1} \nu_x^A h_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ h_{x-1} \log \left( \mu_{x-1} h_{x-1}/\mu_{x-1} \nu_x^A h_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ \nu_x^A \left( h_{x-1} \log \left( g_{x-1}/\nu_x^A g_{x-1} \right) \right) | \sigma_O \right]. \tag{4.22}
\]

From the variational principle (2.6) it follows that

\[
\nu_x^A \left[ h_{x-1} \log \left( g_{x-1}/\nu_x^A g_{x-1} \right) \right] \leq \operatorname{Ent}_x(h_{x-1} | \sigma_O, A), \tag{4.23}
\]

which combined with (4.22) proves (4.20). This completes the proof of part 2.

We use a similar argument for part 1. Let \( \psi_x = \nu(f \mid \sigma, A_{x+1}, \ldots, A_n) \), so that \( \psi_0 = f \) and \( \psi_n = \nu(f \mid \sigma) \). Notice that \( \nu_x[\psi_{x-1} | \sigma] = \psi_x \). Therefore, as in (4.15),

\[
\operatorname{Ent}_\nu(f \mid \sigma) = \sum_{x=1}^n \nu \left[ \operatorname{Ent}_x(\psi_{x-1} | \sigma) | \sigma \right].
\]

Taking expectations with respect to \( \nu(\cdot | \sigma_O) \) we see that it is sufficient to show that, for all \( x \in E \),

\[
\nu \left[ \operatorname{Ent}_x(g_{x-1} | \sigma) | \sigma_O \right] \leq \nu \left[ \operatorname{Ent}_x(\psi_{x-1} | \sigma) | \sigma_O \right]. \tag{4.24}
\]

Introducing the measures \( \mu_k = \otimes_{x=1}^k \nu_x(\cdot | \sigma_O) \), \( \mu_k^A = \otimes_{x=1}^k \nu_x(\cdot | \sigma) \), and \( \nu_x^\sigma = \nu_x(\cdot | \sigma) \), we have \( g_x = \mu_x[f], \psi_x = \mu_x^A[f], \) and \( g_x = \mu_x[\psi_x] \). As in (4.21), we have the commutation relation

\[
\nu_x^\sigma g_{x-1} = \nu_x^\sigma \mu_{x-1} \psi_{x-1} = \mu_{x-1} \nu_x^\sigma \psi_{x-1}. \]

Therefore, as in (4.22)-(4.23) we obtain

\[
\nu \left[ \operatorname{Ent}_x(g_{x-1} | \sigma) | \sigma_O \right] = \nu \left[ g_{x-1} \log \left( g_{x-1}/\nu_x^\sigma g_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ \mu_{x-1} \psi_{x-1} \log \left( \mu_{x-1} \psi_{x-1}/\mu_{x-1} \nu_x^\sigma \psi_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ \psi_{x-1} \log \left( \mu_{x-1} \psi_{x-1}/\mu_{x-1} \nu_x^\sigma \psi_{x-1} \right) | \sigma_O \right] \\
= \nu \left[ \nu_x^\sigma \left( \psi_{x-1} \log \left( g_{x-1}/\nu_x^\sigma g_{x-1} \right) \right) | \sigma_O \right] \\
\leq \nu \left[ \operatorname{Ent}_x(\psi_{x-1} | \sigma) | \sigma_O \right].
\]

This proves (4.24) and completes the proof of part 1. \( \square \)

5 Entropy decay for dynamics in the joint space

In this section we study the implications of our factorization of entropy with respect to the joint measure \( \nu \) (i.e., Theorem 4.2) for dynamics on the joint space.
5.1 Swendsen-Wang in the joint space

We consider several natural variants of the SW dynamics on the joint space $\Omega_j$. Recall that $K$ denotes the $|\Omega_j| \times |\Omega_j|$ stochastic matrix corresponding to re-sampling the spins of a joint configuration given the edges, and similarly $T$ is the stochastic matrix corresponding to re-sampling the edges given the spins. Specifically,

$$K((\sigma, A), (\tau, B)) = 1(A = B)\nu(\tau \mid A)$$

$$T((\sigma, A), (\tau, B)) = 1(\sigma = \tau)\nu(B \mid \sigma).$$

Note that $T = T^* = T^2$ and $K = K^* = K^2$; i.e., $K$ and $T$ are self-adjoint idempotent operators.

As mentioned in the introduction, the Markov chains with transition matrices $KT$ and $TK$ are natural variants of the SW dynamics in the joint space. In the terminology of [17], they are the Markov chains in the joint space corresponding to the two-component Gibbs sampler. The chains with transition matrices $\frac{1}{2}(K + T)$, $KTK$ and $TKT$ are also of interest as the additive and multiplicative reversiblizations of $KT$ and $TK$. We now prove Theorem 1.8 from the introduction, which states that, under SSM, all these dynamics satisfy entropy decay with respect to $\nu$ and hence have $O(\log n)$ mixing time.

First we state the following lemma, which is proved later and will be useful in several of our proofs, including that of Theorem 1.8.

Lemma 5.1. Let $S$ and $S'$ be two idempotent stochastic matrices reversible with respect to a distribution $\pi$ over $\Gamma$, and let $Q = \frac{S + S'}{2}$. Suppose there exists $\delta \in (0, 1)$ such that, for any positive function $f : \Gamma \mapsto \mathbb{R}$, we have $\text{Ent}_\pi(Qf) \leq (1 - \delta)\text{Ent}_\pi(f)$. Then $\text{Ent}_\pi(SS'f) \leq (1 - \delta)\text{Ent}_\pi(f)$ and $\text{Ent}_\pi(S'Sf) \leq (1 - \delta)\text{Ent}_\pi(f)$.

We are now ready to prove Theorem 1.8.

Proof of Theorem 1.8. Let us consider first the case when $P = \frac{K + T}{2}$. Since $P = P^*$, from Lemma 2.4 and Remark 2.5 it is sufficient to prove that, for all positive functions $f : \Omega_i \mapsto \mathbb{R}_+$ with $\mu[f] = 1$,

$$\text{Ent}_\nu Pf \leq (1 - \delta)\text{Ent}_\nu f.$$

The convexity of the function $x \log x$ implies

$$Pf \log Pf \leq \frac{1}{2}Kf \log(Kf) + \frac{1}{2}Tf \log(Tf). \quad (5.1)$$

If $\nu[f] = 1$, then $\nu[Pf] = \nu[Kf] = \nu[Tf] = 1$, and therefore taking expectations with respect to $\nu$ in (5.1) we obtain

$$\text{Ent}_\nu Pf \leq \frac{1}{2} [\text{Ent}_\nu(Kf) + \text{Ent}_\nu(Tf)]. \quad (5.2)$$

Noting that $Kf(\sigma, A) = \nu(f \mid A)$ and $Tf(\sigma, A) = \nu(f \mid \sigma)$, the decompositions in (2.3) and (2.4) imply

$$\text{Ent}_\nu f = \text{Ent}_\nu(Kf) + \nu [\text{Ent}_\nu(f \mid A)] = \text{Ent}_\nu(Tf) + \nu [\text{Ent}_\nu(f \mid \sigma)].$$

Hence, (5.2) becomes

$$\text{Ent}_\nu Pf \leq \text{Ent}_\nu f - \frac{1}{2} \nu [\text{Ent}_\nu(f \mid A) + \text{Ent}_\nu(f \mid \sigma)]. \quad (5.3)$$
Theorem 1.7 now implies
\[ \text{Ent}_\nu(Pf) \leq (1 - \delta)\text{Ent}_\nu(f), \]
with \( \delta = 1/2C. \) This proves the theorem for the case when \( P = \frac{K + T}{2}. \) The result for \( KT \) and \( TK \) follows from Lemma 5.1, by recalling that \( K^2 = K = K^*, T^2 = T = T^*, \) and noting that \( (KT)^* = TK \) and \( (TK)^* = KT. \) Finally, the cases \( P = KTK, P = TKT \) follow from the cases \( P = KT \) and \( P = TK \) with the observation that, by (5.6), \( \text{Ent}_\nu(KTKf) \leq \text{Ent}_\nu(TKf) \) and \( \text{Ent}_\nu(TKf) \leq \text{Ent}_\nu(KTf). \)

Finally, we go back and supply the missing proof of Lemma 5.1.

\textbf{Proof of Lemma 5.1.} Let us first show that
\[ \text{Ent}_\pi(Sf) \leq \text{Ent}_\pi(\tilde{S}f), \quad (5.4) \]
where \( \tilde{S} = \frac{S + I}{2} \) is a lazy version of \( S; \) \( I \) denotes the identity matrix. To this end, define \( U_n = \left[ \frac{1}{n}(S + I) \right]^n. \) Then we have \( U_1 = \tilde{S} \) and \( U_n = \left( \frac{2n^2 - 1}{2n} \right)^n S \) as \( n \to \infty. \) Therefore, (5.4) follows if we prove that for all \( n \geq 1 \)
\[ \text{Ent}_\pi(U_{n+1}f) \leq \text{Ent}_\pi(U_n f). \quad (5.5) \]
On the other hand, if \( U \) is any stochastic matrix with stationary distribution \( \pi, \) then for any function \( f : \Gamma \to \mathbb{R}_+ \) with \( \pi[f] = 1 \) we have \( \pi[UF] = 1. \) Hence, \( \text{Ent}_\pi(U f) = \pi[(U f) \log(U f)]. \) Since \( U \) is a stochastic matrix, the convexity of the function \( x \log x \) implies \( (U f) \log(U f) \leq U(f \log f), \) and so
\[ \text{Ent}_\pi(U f) \leq \pi[U(f \log f)] = \pi[f \log f] = \text{Ent}_\pi(f), \quad (5.6) \]
Since \( U_{n+1}f = U_1U_n f, \) applying (5.6) with \( f \) replaced by \( U_n f \) and with \( U = U_1 \) proves (5.5) and (5.4). We note that since \( (S')^2 = S', \)
\[ \tilde{S}S' = \frac{1}{2}(S + S')S' = QS'. \]
Applying (5.4) with \( f \) replaced by \( S'S \) we obtain
\[ \text{Ent}_\pi(SS'f) \leq \text{Ent}_\pi(\tilde{S}S'f) = \text{Ent}_\pi(QS'f) \leq (1 - \delta)\text{Ent}_\pi(S'f) \leq (1 - \delta)\text{Ent}_\pi(f), \]
where the second inequality follows from the assumption that \( Q \) contracts entropy for any function and the last one follows again from (5.6). This completes the proof for \( SS'. \) The same argument with \( S \) and \( S' \) exchanged applies for \( S'S \) and we are done.

\textbf{5.2 The local dynamics in the joint space}

In this section, we show that the natural Glauber dynamics for the joint measure \( \nu \) mixes in time \( O(n \log n) \) whenever the spin marginal of \( \nu \) satisfies the SSM property. In particular, we prove Theorem 1.9 from the introduction.

For any \( v \in V, e \in E, \) let \( Q_v \) denote the stochastic matrix corresponding to the single heat-bath update at vertex \( v, \) and let \( W_e \) denote the stochastic matrix for the single heat-bath update at the edge \( e. \) Then the transition matrix \( P_{\text{local}} \) of the Glauber dynamics in the joint space is given by
\[ P_{\text{local}} = \frac{1}{2|V|} \sum_{v \in V} Q_v + \frac{1}{2|E|} \sum_{e \in E} W_e. \quad (5.7) \]
As mentioned in the introduction, the heat-bath updates in the joint space are quite simple. For a vertex \( v \in V \), the heat-bath update at \( v \) assigns a new spin to \( v \) chosen u.a.r. from \( \{1, \ldots, q\} \), provided \( v \) is isolated (i.e., there are no edges incident to \( v \) in the edge configuration); otherwise, the spin at \( v \) does not change. On the other hand, the heat-bath update at \( e \in E \) updates the state of \( e \) only if it is monochromatic in the spin configuration; if this is the case, the new state of \( e \) corresponds to a Bernoulli(\( p \)) random variable. We note that \( Q_v \) and \( W_e \) are reversible with respect to \( \nu \). Moreover, they are projection operators in \( L^2(\Omega, \nu) \); that is, \( Q_v^2 = Q_v = Q_v^* \) and \( W_e^2 = W_e = W_e^* \).

**Proof of Theorem 1.9.** First note that since \( Q_v \) and \( W_e \) are reversible with respect to \( \nu \), so is \( P_{\text{local}} \) and by Lemma 2.4 and Remark 2.5 it is sufficient for us to establish that

\[
\text{Ent}_\nu(P_{\text{local}} f) \leq (1 - \delta/n) \text{Ent}_\nu(f)
\]

(5.8)

for all positive functions \( f : \Omega \to \mathbb{R}_+ \) such that \( \nu[f] = 1 \). Here \( \delta > 0 \) is a constant independent of \( n \) and the admissible boundary condition.

By the convexity of the function \( x \log x \), reasoning as in (5.2), we can write

\[
\text{Ent}_\nu(P_{\text{local}} f) \leq \frac{1}{2|V|} \sum_{v \in V} \text{Ent}_\nu(Q_v f) + \frac{1}{2|E|} \sum_{e \in E} \text{Ent}_\nu(W_e f).
\]

Let \( \sigma_{V \setminus v} \) (resp., \( A_{E \setminus e} \)) denote the spin (resp., edge) configuration excluding \( v \) (resp., excluding \( e \)). Since \( Q_v f(\sigma, A) = \nu(f \mid \sigma_{V \setminus \{v\}}, A) \) and \( W_e f(\sigma, A) = \nu(f \mid \sigma, A_{E \setminus \{e\}}) \), from the decompositions of entropy in (2.3) and (2.4) we obtain

\[
\text{Ent}_\nu(Q_v f) = \text{Ent}_\nu(f) - \nu \left[ \text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A) \right];
\]

\[
\text{Ent}_\nu(W_e f) = \text{Ent}_\nu(f) - \nu \left[ \text{Ent}_\nu(f \mid \sigma, A_{E \setminus \{e\}}) \right].
\]

Therefore,

\[
\text{Ent}_\nu(P_{\text{local}} f) \leq \text{Ent}_\nu(f) - \frac{1}{2|V|} \sum_{v \in V} \nu \left[ \text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A) \right] - \frac{1}{2|E|} \sum_{e \in E} \nu \left[ \text{Ent}_\nu(f \mid \sigma, A_{E \setminus \{e\}}) \right].
\]

We show next that there exists a constant \( C > 0 \) such that

\[
\text{Ent}_\nu(f) \leq C \sum_{v \in V} \nu \left[ \text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A) \right] + C \sum_{e \in E} \nu \left[ \text{Ent}_\nu(f \mid \sigma, A_{E \setminus \{e\}}) \right].
\]

(5.9)

The desired estimate (5.8) then follows from the fact that \( |E| = O(|V|) = O(n) \).

To establish (5.9), note that from Lemma 4.6 we know that, for some constant \( C_1 > 0 \),

\[
\text{Ent}_\nu(f) \leq C_1 \nu \left[ \text{Ent}_\nu(f \mid \sigma_E) + \text{Ent}_\nu(f \mid \sigma_O) \right],
\]

(5.10)

where we recall that \( E \subset V \) and \( O \subset V \) are the even and odd sub-lattices, respectively. Since \( \nu(\cdot \mid \sigma_O) = \otimes_{v \in E} \nu_v(\cdot \mid \sigma_O) \) (see (4.8)), the standard tensorization of entropy for product measures (see, e.g., [3]) implies

\[
\text{Ent}_\nu(f \mid \sigma_O) \leq \sum_{v \in E} \nu \left[ \text{Ent}_\nu(f \mid \sigma_O) \mid \sigma_O \right],
\]

(5.11)
where as before we use $\text{Ent}_\nu(f \mid \sigma_V)$ for the entropy with respect to $\nu_v(\cdot \mid \sigma_V)$. From Lemma 4.8 we see that

$$\text{Ent}_\nu(f \mid \sigma_V) \leq C_1 \sum_{v \in E} \nu [\text{Ent}_\nu(f \mid \sigma_V, A) + \text{Ent}_\nu(f \mid \sigma) \mid \sigma_V], \quad (5.12)$$

for some constant $C_1 > 0$.

For $v \in E$, the distribution of the spin $\sigma_v$ given $\sigma_V$ and $A$ is the same as the distribution of $\sigma_v$ given $\sigma_{V \setminus \{v\}}$ and $A$; that is, $\nu_v(\cdot \mid \sigma_V, A) = \nu(\cdot \mid \sigma_{V \setminus \{v\}}, A)$. Therefore we may write

$$\text{Ent}_\nu(f \mid \sigma_V, A) = \text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A). \quad (5.13)$$

Let us also observe that, for every $v \in E$, $\nu_v(\cdot \mid \sigma)$ is a product measure on $A_v = \{A_{vw}, \{w, v\} \in \mathbb{E}\}$, and the entropy appearing on the right hand side above is simply the entropy of $A_{vw}$ once every other spin or edge variable has been fixed. Therefore, (5.14) is again the standard tensorization statement for product measures. In conclusion, we have shown that

$$\text{Ent}_\nu(f \mid \sigma_V) \leq C_1 \sum_{v \in E} \nu [\text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A) \mid \sigma_V] + C_1 \sum_{e \in \mathbb{E}} \nu [\text{Ent}_\nu(f \mid \sigma, A_{\mathbb{E} \setminus e}) \mid \sigma_V], \quad (5.15)$$

where the second sum is now over the set of all edges $\mathbb{E}$. The same estimate can be obtained with the role of even and odd sites reversed:

$$\text{Ent}_\nu(f \mid \sigma_V) \leq C_1 \sum_{v \in O} \nu [\text{Ent}_\nu(f \mid \sigma_{V \setminus \{v\}}, A) \mid \sigma_O] + C_1 \sum_{e \in \mathbb{E}} \nu [\text{Ent}_\nu(f \mid \sigma, A_{\mathbb{E} \setminus e}) \mid \sigma_O]. \quad (5.16)$$

Taking expectations with respect to $\nu$ and summing (5.15) and (5.16), from (5.10) we obtain (5.9) which finishes the proof.

**Remark 5.2.** By taking $f$ that depends only on spins, we derive as a corollary of Theorem 1.9 entropy decay for the Potts model Glauber dynamics (up to a constant laziness factor to account for the probability of a site being isolated); similarly, taking $f$ that depends only on edges, we obtain entropy decay for the corresponding Glauber dynamics for the random-cluster model. While entropy decay was previously known for the Potts Glauber dynamics under SSM [15], the same statement for the random-cluster dynamics appears to be a new result. (Note in particular that entropy decay does not follow from the mixing time results for this dynamics in [9].)

### 6 Entropy decay for the alternating scan dynamics

We consider next entropy decay for the alternating scan dynamics for the Potts model. In particular, we prove Theorem 1.6 from the introduction. Recall that $P_E$ is the transition matrix corresponding to equilibration on the even sites conditional on the spins of the odd sites, and $P_O$ is the same for the odd sites. We consider the alternating scan chains defined by the stochastic matrices $S_{EO} = P_E P_O$ and $S_{OE} = P_O P_E$. Note that $P_E$ and $P_O$ do not commute (unless $\beta = 0$), and thus $S_{EO}, S_{OE}$ are not reversible with respect to $\mu$. Non-reversibility is often a serious obstacle for the analysis of Markov chains, and while in practice $S_{EO}$ and $S_{OE}$ are the most widely used versions of alternating scan dynamics, previous theoretical analyses had focused on simpler reversible versions, namely the multiplicative reversibilizations $P_O P_E P_O$ and $P_E P_O P_E$ [5, 27], and the additive reversibilization $P_E + P_O$ [13].
Proof of Theorem 1.6. We will show that the discrete entropy contraction in (2.12) holds for $S_{EO}$ and $S_{OE}$ for any positive function $f : \Omega \to \mathbb{R}$ such that $\mu [f] = 1$. The mixing time bounds then follow from Lemma 2.4 and the fact that $S_{EO}^* = S_{OE}$ and $S_{OE}^* = S_{EO}$. In view of Lemma 5.1, it is sufficient for us to establish (2.12) for $P := \frac{P_E + P_O}{2}$. The convexity of the function $x \log x$ implies the pointwise bound

$$
(Pf) \log (Pf) \leq \frac{1}{2} (P_Ef) \log (P_Ef) + \frac{1}{2} (P_Of) \log (P_Of).
$$

From this and the fact that $\mu [Pf] = 1$ we get

$$
\text{Ent}_\mu (Pf) = \mu [(Pf) \log (Pf)] \leq \frac{1}{2} \left[ \text{Ent}_\mu (P_Ef) + \text{Ent}_\mu (P_Of) \right].
$$

(6.1)

Note that $P_E$ and $P_O$ are the orthogonal projections in $L^2 (\Omega, \mu)$ such that $P_Ef = \mu (f \mid \sigma_O)$ and $P_Of = \mu (f \mid \sigma_E)$. Therefore,

$$
\begin{align*}
\text{Ent}_\mu (f) &= \text{Ent}_\mu (f \mid \sigma_O) + \mu \left[ \text{Ent}_\mu (f \mid \sigma_O) \right] = \text{Ent}_\mu (P_Ef) + \mu \left[ \text{Ent}_\mu (f \mid \sigma_O) \right]; \\
\text{Ent}_\mu (f) &= \text{Ent}_\mu (f \mid \sigma_E) + \mu \left[ \text{Ent}_\mu (f \mid \sigma_E) \right] = \text{Ent}_\mu (P_Of) + \mu \left[ \text{Ent}_\mu (f \mid \sigma_E) \right],
\end{align*}
$$

and we see that (6.1) is equivalent to

$$
\text{Ent}_\mu (Pf) \leq \text{Ent}_\mu (f) - \frac{1}{2} \mu \left[ \text{Ent}_\mu (f \mid \sigma_O) + \text{Ent}_\mu (f \mid \sigma_E) \right].
$$

We may now apply Theorem 4.3 which implies that, when SSM holds,

$$
\text{Ent}_\mu (Pf) \leq (1 - \delta) \text{Ent}_\mu (f),
$$

(6.2)

for a suitable constant $\delta \in (0, 1)$. This establishes (2.12) for $P = \frac{P_E + P_O}{2}$. Since $P_E^2 = P_E = P_E^*$, $P_O^2 = P_O = P_O^*$, and $(P_E P_O)^* = P_O P_E$, $(P_O P_E)^* = P_E P_O$, the remainder of the result follows from Lemma 5.1.

\[\square\]

7 Random-cluster dynamics

In this section we study the implications of our results for the dynamics of the random-cluster model for both the high and low temperatures regimes. This allows us to derive Theorem 1.3 from the introduction using a comparison mechanism we establish in Section 7.2.

The random-cluster model on $G = (V, E)$ with parameters $p \in (0, 1)$ and $q > 0$ assigns to each $A \subseteq E$ a probability

$$
\varrho (A) = \varrho_{G,p,q} (A) = \frac{1}{Z_{RC}} p^{|A|(1-p)^{|E|-|A|}} q^{c(A)},
$$

(7.1)

where $c(A)$ is the number of connected components in $(V, A)$ and $Z_{RC}$ is the corresponding partition function. The random-cluster model was first introduced by Fortuin and Kasteleyn [22] as a unifying framework for random graphs, spin systems and electrical networks; see the book [26] for extensive background.

A boundary condition for the random-cluster model is a partition $\xi = \{ \xi_1, \xi_2, \ldots \}$ of the internal boundary $\partial V$ of $V$ such that all vertices in each $\xi_i$ are constrained to be in the same connected component of any configuration $A$. (We can think of the vertices in $\xi_i$ as being connected through a configuration in $V^c$.) These connections are considered in the counting of the connected components in (7.1); i.e., $c(A)$ becomes $c(A, \xi)$ (see, e.g., [7, 26]).
The distribution $g$ with a free boundary condition (i.e., every element of $\xi$ is a single vertex) corresponds to the edge marginal of the joint measure also with free boundary condition (1.3); that is, $g(A) = \sum_{A \subseteq M(\sigma)} \nu(\sigma, A)$ and $Z_{\mathbb{R}C} = Z_1$; see, e.g., [21, 26]. The wired boundary condition corresponds to the case when all vertices of $\partial V$ are connected by the boundary condition (i.e., $\xi = \{\partial V\}$). More generally, if $(\psi, \varphi)$ is an admissible boundary condition for the joint space (see Definition 4.1), we have

$$g^{\psi, \varphi}(A) = \sum_{\sigma: A \subseteq M(\sigma)} \nu^{\psi, \varphi}(\sigma, A) = \frac{1}{Z_{\psi, \varphi}} p^{|A|}(1 - p)^{|E| - |A|} q^{c(A) - c_0(A)} 1(A \sim \psi) 1(A \sim \varphi),$$  

where $A \sim \psi$ means that $A$ does not connect vertices of $V_0$ with different colors in $\psi$, $A \sim \varphi$ that $A$ and $\varphi$ agree on the edges in $E_0$ and $c_0(A)$ denotes the number of connected components that intersect $V_0 \subseteq \partial V$; see Figure 7.1 for some admissible boundary conditions.

As an example, consider the admissible boundary condition that is obtained by taking $V_0 = \partial V$, $E_0 = \partial E$, with $\psi = i$ for some $i \in [q]$ (i.e., the monochromatic spin boundary condition) and $\varphi = 1$; see Figure 7.1(c). In this case, $g^{\psi, \varphi}$ is the random-cluster measure on the cube $R = \{1, \ldots, \ell - 1\}^d \subset V$ with wired boundary condition. On the other hand, the marginal on the spins is the Potts measure on $R$ with the “all $i$” monochromatic boundary condition.

Another relevant random-cluster boundary condition is the one obtained by adding to the random-cluster space the edges “sticking in” from $\partial V$. Namely, let $E_1 \subset \partial E$ be the set of edges with exactly one endpoint in $\partial V$, and take the monochromatic boundary condition $\psi = i$ and the wired edge boundary condition on $E_0 = \partial E \setminus E_1$. The marginal on edges is the random-cluster distribution measure on $(V, E \setminus E_0)$ with wired boundary condition on $\partial V$, while the spin marginal is the Potts measure on $R$ with the “all $i$” boundary condition on $\partial V$; see Figure 7.1(d).

Reasoning in this way one can obtain, as the edge marginal of the joint measure with an admissible boundary condition, any random-cluster measure with a boundary condition where the vertices in the boundary are either free or wired into a single component, simply by fixing monochromatic spins on that component and fixing an edge configuration realizing the wiring of that component.

**Planar duality.** A useful tool in two dimensions is planar duality. Let $G_d = (V_d, E_d)$ denote the planar dual of $G = (V, E)$, where $V = \{0, \ldots, \ell\} \times \{0, \ldots, \ell\}$ is a square region of $\mathbb{Z}^2$. That is, $V_d$...
corresponds to the set of faces of \( V \), and for each \( e \in \mathcal{E} \), there is a dual edge \( e_d \in \mathcal{E}_d \) connecting the two faces bordering \( e \). The random-cluster distribution (7.1) satisfies \( \varrho_{G,p,q}(A) = \varrho_{G_d,p_d,q}(A_d) \), where \( A_d \) is the dual configuration to \( A \subseteq \mathcal{E} \); i.e., \( e_d \in A_d \) iff \( e \not\in A \), and

\[
p_d = \frac{q(1-p)}{q(1-p) + p}.
\]

The self-dual point (i.e., the value of \( p \) such that \( p = p_d \)) corresponds to the critical threshold \( p_c(q) = 1 - \exp(-\beta c(q)) \).

Since \( V_d \) is not a subset of \( \mathbb{Z}^2 \), it is convenient to consider the graph \( \hat{G}_d = (\hat{V}_d, \hat{E}_d) \) and identify all boundary vertices of \( \hat{V}_d \) with the vertex of \( G_d \) corresponding to its external face. Then, \( \varrho^1_{G_d,p_d,q}(A) = \varrho^0_{G_d,p_d,q}(A_d) \) and \( \varrho^0_{G_d,p_d,q}(A) = \varrho^1_{G_d,p_d,q}(A_d) \), where the 0 and 1 superscripts denote the free and and wired boundary conditions respectively (see Section 6.1 in [26] for a detailed discussion).

Observe also that both random cluster measures \( \varrho^1_{G_d,p_d,q} \) and \( \varrho^0_{G_d,p_d,q} \) on \( \hat{G}_d \) can be obtained as marginals of the joint measure in a square region of \( \mathbb{Z}^2 \) with a monochromatic admissible boundary condition as described above.

7.1 SW dynamics for the random-cluster model

Our first result concerns the SW dynamics for the random-cluster model. In this variant of the SW dynamics, given an edge configuration \( A \), we assign spins to the connected components of \( A \) uniformly at random to obtain a joint configuration, and then update the edge configuration by percolating on the monochromatic edges with probability \( p \). The transition matrix \( \widetilde{P}_{sw} \) of this chain satisfies

\[
\widetilde{P}_{sw}(A, B) = \sum_{\sigma: \mathcal{A} \subseteq \mathcal{M}(\sigma)} \nu(\sigma \mid A)\nu(\sigma \mid B);
\]

\( \widetilde{P}_{sw} \) is reversible with respect to \( \varrho \); see, e.g., [21, 49]. The following lemma follows from Theorem 1.8.

**Lemma 7.1.** Let \( \nu := \nu^\psi\varphi \) be the joint distribution with an admissible boundary condition (\( \psi \), \( \varphi \)). If \( q \) and \( \beta = \ln\left(\frac{1}{1-p}\right) \) are such that SSM holds, then the SW dynamics on random-cluster configurations with boundary conditions inherited from (\( \psi \), \( \varphi \)) satisfies the discrete time entropy decay with rate \( \delta \), and its mixing time is bounded by \( O(\log n) \).

**Proof.** If \( f \) depends only on the edge configuration, then

\[
\widetilde{P}_{sw}f(A) = \nu[f \mid \sigma \mid A] = TKf(\sigma, A).
\]

(7.3)

Here and below, with slight abuse of notation, if a function \( f \) on the joint space depends only on the edge configuration, we again write \( f \) for the corresponding (projection) function on edges. Therefore, we have \( \text{Ent}_\varrho(\widetilde{P}_{sw}f) = \text{Ent}_\varrho(TKf) \). More precisely, for any \( f \geq 0 \) depending only on the edge configuration, and such that \( \varrho[f] = \nu[f] = 1 \), one has

\[
\text{Ent}_\varrho(\widetilde{P}_{sw}f) = \varrho[\widetilde{P}_{sw}f \log(\widetilde{P}_{sw}f)] = \nu[(TKf) \log(TKf)] = \text{Ent}_\varrho(TKf).
\]

Theorem 1.8 says that, for any function \( f \) in the joint space, one has

\[
\text{Ent}_\varrho[TKf] \leq (1 - \delta)\text{Ent}_\varrho(f).
\]

In particular, for our \( f \),

\[
\text{Ent}_\varrho(\widetilde{P}_{sw}f) \leq (1 - \delta)\text{Ent}_\varrho(f) = (1 - \delta)\text{Ent}_\varrho(f).
\]

This is the desired discrete time entropy decay for \( \widetilde{P}_{sw} \) in the edge space. \( \Box \)
Remark 7.2. The same argument in the previous proof applies to the spin dynamics. In particular, if \( g \) is a function depending only on the spin configuration, then \( P_{sw} g(\sigma) = K T g(\sigma, A) \). Repeating the previous steps with \( K T \) in place of \( T K \) one has discrete time entropy decay with rate \( \delta \) for the SW dynamics on spin configurations. This provides an alternative view of the proof of Theorem 1.1 as a corollary of Theorem 1.8 for the joint space.

In \( \mathbb{Z}^2 \), we can take advantage of self-duality of the random-cluster model to obtain bounds for the SW dynamics in the low temperature regime.

Theorem 7.3. In an \( n \)-vertex square region of \( \mathbb{Z}^2 \) with free or wired boundary conditions, for all integer \( q \geq 2 \) and all \( p > p_c(q) \), there exists a constant \( \delta > 0 \) such that for all functions \( f : \{0, 1\}^E \to \mathbb{R}_+ \)

\[
\text{Ent}_\theta(P_{sw} f) \leq \left( 1 - \frac{\delta}{n} \right) \text{Ent}_\theta(f).
\]

In particular, the mixing time of the SW dynamics on random-cluster configurations satisfies \( T_{\text{mix}}(\tilde{P}_{sw}) = O(n \log n) \).

Let \( G = (V, E) \) where \( V \) is an \( n \)-vertex square region of \( \mathbb{Z}^2 \). Let \( g := g^\theta \) where \( \theta \in \{0, 1\} \) and let \( P_{hb} \) be the transition matrix of the heat-bath Glauber dynamics on \( G \). This is the standard Markov chain that, from a random-cluster configuration \( A_t \subseteq E \), transitions to a new configuration \( A_{t+1} \subseteq E \) as follows:

1. choose an edge \( e \in E \) uniformly at random;
2. let \( A_{t+1} = A_t \cup \{e\} \) with probability

\[
\frac{g(A_t \cup \{e\})}{g(A_t \cup \{e\}) + g(A_t \setminus \{e\})} = \begin{cases} 
\frac{p}{q(1-p)+p} & \text{if } e \text{ is a "cut-edge" in } (V, A_t); \\
\frac{p}{q(1-p)+p} & \text{otherwise}; 
\end{cases}
\]

3. otherwise, let \( A_{t+1} = A_t \setminus \{e\} \).

We say \( e \) is a cut-edge in \( (V, A_t) \) if the number of connected components in \( A_t \cup \{e\} \) and \( A_t \setminus \{e\} \) differ. \( P_{hb} \) is (by design) reversible with respect to \( g \). It is also straightforward to check that with the free (resp., wired) boundary condition and parameters \( p \) and \( q \), for any pair of configurations \( A \) and \( B \), we have \( P_{hb}(A, B) = P_{hb}^d(A_d, B_d) \), where \( P_{hb}^d \) denotes the transition matrix of the heat-bath chain on \( \hat{G}_d \) with wired (resp., free) boundary condition and parameters \( p_d \) and \( q \).

Theorem 7.3 follows from the following two results.

Lemma 7.4. There exists a constant \( c > 0 \) such that, for every function \( f : \{0, 1\}^E \to \mathbb{R} \),

\[
\mathcal{D}_{P_{sw}}(f, f) \geq c \cdot \mathcal{D}_{P_{hb}}(f, f).
\]

Lemma 7.5. For all integer \( q \geq 2 \) and all \( p > p_c(q) \), there exists a constant \( \delta > 0 \) such that, for every function \( f : \{0, 1\}^E \to \mathbb{R}_+ \),

\[
\mathcal{D}_{P_{hb}}(\sqrt{f}, \sqrt{f}) \geq \frac{\delta}{n} \cdot \text{Ent}_\theta(f).
\]
Proof of Theorem 7.3. Lemmas 7.4 and 7.5 imply

\[ \mathcal{D}_{\tilde{P}_{sw}}(\sqrt{f}, \sqrt{f}) \geq \frac{c \delta}{n} \cdot \text{Ent}_{\varrho}(f). \]  

(7.4)

In words, this says that the SW dynamics on random-cluster configurations when \( p > p_c(q) \) satisfies a standard log-Sobolev inequality with constant \( \frac{c \delta}{n} \). Since \( \tilde{P}_{sw} = \tilde{P}_{sw}^* \), Lemma 2.8 shows that (7.4) implies the entropy decay bound

\[ \text{Ent}_{\varrho}(\tilde{P}_{sw} f) \leq \left( 1 - \frac{\delta c}{n} \right) \text{Ent}_{\varrho}(f), \]

and the mixing time bound follows from Lemma 2.4 and Remark 2.5.

It remains to prove Lemmas 7.4 and 7.5. We note that a version of the comparison inequality in Lemma 7.4 was proved in [49] (see Theorem 4.8 there), but it is stated for the spectral gap under the free boundary condition.

In both of these proofs, we consider the single-bond variant of the Glauber dynamics. In one step of this chain every connected component is assigned a spin from \([q]\) uniformly at random; a random edge \( e \) is then chosen and if the endpoints of \( e \) are monochromatic, then the edge is added to the configuration with probability \( p \) and deleted otherwise. The state of \( e \) does not change if its endpoints are bi-chromatic. Note that this chain is the projection onto edges of the local dynamics on the joint space, see (5.7); in particular, the update at the edge \( e \) corresponds to \( W_e \). Let \( P_{sb} \) denote the transition matrix of the single bond dynamics, which is reversible with respect to \( \varrho \).

The Dirichlet form associated to this chain satisfies

\[ \mathcal{D}_{P_{sb}}(f, f) = \langle (I - P_{sb}) f, f \rangle_{\varrho} = \varrho \left[ ((I - P_{sb}) f) \cdot f \right] = \frac{1}{|E|} \sum_{e \in E} \nu \left[ \text{Var}_{\nu}(f | \sigma, A_{E\setminus e}) \right] \]  

(7.5)

since

\[ P_{sb} f(A) = \frac{1}{|E|} \sum_{e \in E} \nu \left[ \nu[f | \sigma, A_{E\setminus e}] \mid A \right], \]

where with a slight abuse of notation (here and below) we use \( f \) also for the “lift” of \( f \) to the joint space.

We note that for some constants \( c_i = c_i(q, p) > 0, i = 1, 2, \)

\[ c_1 P_{sb}(A, B) \leq P_{HB}(A, B) \leq c_2 P_{sb}(A, B) \]

for all random-cluster configurations \( A, B \). Therefore the same bounds apply to the Dirichlet forms:

\[ c_1 \mathcal{D}_{P_{sb}}(f, f) \leq \mathcal{D}_{P_{fb}}(f, f) \leq c_2 \mathcal{D}_{P_{sb}}(f, f), \]  

(7.6)

for any function \( f : \{0, 1\}^E \mapsto \mathbb{R} \).

Proof of Lemma 7.4. The Dirichlet form associated with \( \tilde{P}_{sw} \) is given by

\[ \mathcal{D}_{\tilde{P}_{sw}}(f, g) = \langle (I - \tilde{P}_{sw}) f, g \rangle_{\varrho} = \varrho \left[ ((I - \tilde{P}_{sw}) f) \cdot g \right], \]

and since \( \tilde{P}_{sw} f(A) = \nu[\nu[f | \sigma] \mid A] \), we obtain

\[ \mathcal{D}_{\tilde{P}_{sw}}(f, f) = \nu [(f - \nu[\nu[f | \sigma] \mid A]) \cdot f] = \nu [(f - \nu[f | \sigma]) \cdot f] = \nu [\text{Var}_{\nu}(f | \sigma)]. \]
Then, for any function \( f \geq 0 \),
\[
\mathcal{D}_{\text{Pm}}(\sqrt{f}, \sqrt{f}) = \nu \left[ \text{Var}_\nu(\sqrt{f} \mid \sigma) \right] \\
\geq \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \nu \left[ \text{Var}_\nu(\sqrt{f} \mid \sigma, A_{\mathcal{E}\setminus e}) \right] \\
= \mathcal{D}_{\text{Pm}}(\sqrt{f}, \sqrt{f}),
\]
where we have used (7.5) and the fact that, for any \( e \in \mathcal{E} \),
\[
\nu \left[ \text{Var}_\nu(\sqrt{f} \mid \sigma) \right] \geq \nu \left[ \text{Var}_\nu(\sqrt{f} \mid \sigma, A_{\mathcal{E}\setminus e}) \right]
\]
by monotonicity of the variance functional. The result then follows from (7.6).

**Proof of Lemma 7.5.** By duality (see discussion at the beginning of the section), we have
\[
\mathcal{D}_{\text{Pm}}(\sqrt{f}, \sqrt{f}) = \mathcal{D}_{\text{Pd}}(\sqrt{f_d}, \sqrt{f_d}),
\]
where \( f_d \) is the function such that \( f_d(A_d) = f(A) \) and \( P_{\text{Hd}}^{\dagger} \) is the transition matrix corresponding to the dual of \( \varrho \).

Thus, if \( \mathcal{D}_{\text{Pm}} \) is at low temperature \((p > p_c(q))\), then \( \mathcal{D}_{\text{Pd}} \) is at high temperature \((p < p_c(q))\).

Moreover, from (7.5) and (7.6),
\[
\mathcal{D}_{\text{Pd}}(\sqrt{f}, \sqrt{f}) \geq c_1 \mathcal{D}_{\text{Pd}}(\sqrt{f}, \sqrt{f}) = \frac{c_1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \nu_d \left[ \text{Var}_{\nu_d}(\sqrt{f} \mid \sigma, A_{\mathcal{E}\setminus e}) \right],
\]
where \( \nu_d \) is the dual joint measure. Specifically, if \( \varrho_d \) is the dual measure of \( \varrho \) (and the stationary distribution of \( P_{\text{Hd}}^{\dagger} \)), \( \nu_d \) is a joint measure whose edge marginal is \( \varrho_d \). Observe that since \( \varrho \) is a random-cluster distribution on the square region \( \mathcal{V} = \{0, \ldots, n\} \times \{0, \ldots, \ell\} \) with free (or wired) boundary condition, \( \varrho_d \) is a distribution over \( \tilde{\mathcal{V}}_d = \{-1, \ldots, \ell\} \times \{-1, \ldots, \ell\} \times \{x, \frac{1}{2}, \frac{1}{2}\} \) with wired (or free) boundary condition. As discussed earlier, in either case there exists a joint measure with an admissible boundary condition whose edge marginal is \( \varrho_d \).

Observe also that, as before, with a slight abuse of notation, we also use \( f \) for the “lift” of \( f \) to the joint space. Now, as in (4.13) we know that for some constant \( C = C(p, q) \), for all \( e \in \mathcal{E} \) and for all \( f \geq 0 \),
\[
\text{Var}_{\nu_d}(\sqrt{f} \mid \sigma, A_{\mathcal{E}\setminus e}) \geq C^{-1} \text{Ent}_{\nu_d}(f \mid \sigma, A_{\mathcal{E}\setminus e}).
\]

Therefore,
\[
\mathcal{D}_{\text{Pd}}(\sqrt{f}, \sqrt{f}) \geq \frac{c_1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \nu_d \left[ \text{Ent}_{\nu_d}(f \mid \sigma, A_{\mathcal{E}\setminus e}) \right].
\]

Since for \( p < p_c(q) \) and \( q \geq 2 \) the SSM property holds, we can use (5.9) to obtain
\[
\sum_{e \in \mathcal{E}} \nu_d \left[ \text{Ent}_{\nu_d}(f \mid \sigma, A_{\mathcal{E}\setminus e}) \right] \geq \delta_1 \text{Ent}_{\nu_d}(f).
\]

Indeed, if \( f \) is a function of edges only then the first term on the right hand side of (5.9) is zero. Moreover for such an \( f \) we have \( \text{Ent}_{\nu_d}(f) = \text{Ent}_{\varrho_d}(f) \). Summarizing, we have proved, for all \( f \geq 0 \),
\[
\mathcal{D}_{\text{Pd}}(\sqrt{f_d}, \sqrt{f_d}) \geq \frac{\delta_2}{n} \text{Ent}_{\varrho_d}(f_d),
\]
for a suitable constant \( \delta_2 > 0 \). The result follows from (7.7) and the fact that \( \text{Ent}_{\varrho_d}(f_d) = \text{Ent}_{\varrho}(f) \). \( \square \)
Remark 7.6. We remark that \((7.10)\) says that the heat-bath Glauber dynamics for the random-cluster model in square regions of \(\mathbb{Z}^2\) with free or wired boundary conditions satisfies the standard log-Sobolev inequality with constant \(\delta/n\) for some \(\delta = \delta(p, q)\) for all \(p \neq p_c(q)\). This bound is optimal up to a multiplicative constant, as can be seen by choosing an appropriate test function.

7.2 Decay for spins from decay for edges and vice versa

We will use Theorem 7.3 to deduce our low temperature results for the SW dynamics on spin configurations. We do so using the following entropy contraction “transfer” result between the spin and edge variants of the SW dynamics. A similar comparison result for the spectral gap was provided by Ullrich \([49]\).

Lemma 7.7. Suppose we know that the SW dynamics on edges with invariant measure \(\nu\), corresponding to an \(n\)-vertex square region \(V\) with some boundary condition, has entropy decay with rate \(\delta\). Then the SW dynamics on spins on \(V\), with any boundary condition inherited from a joint measure \(\nu\) whose marginal on edges equals \(\nu\), satisfies the same entropy decay (asymptotically) and has the same mixing time bound \(T_{\text{mix}} = O(\delta^{-1} \log n)\). The same applies with the roles of spins and edges reversed.

Proof. The assumption on \(\nu\) says that
\[
\text{Ent}_\nu(\tilde{P}_{SW} g) \leq (1 - \delta) \text{Ent}_\nu(g),
\]
(7.11)
for any function \(g = g(A), A \subset E\). Recalling \((7.3)\) we see that \((7.11)\) can be rewritten as
\[
\text{Ent}_\nu(TK g) \leq (1 - \delta) \text{Ent}_\nu(g),
\]
(7.12)
for any \(g = g(A)\) and any joint measure \(\nu\) such that the marginal on edges equals \(\nu\). Now, let \(f = f(\sigma)\) be any function depending only on the spin configuration. Since \(g = Tf\) depends only on the edge configuration, we have
\[
\text{Ent}_\nu(TK T f) \leq (1 - \delta) \text{Ent}_\nu(Tf).
\]
(7.13)
If we apply \((7.13)\) with \(f\) replaced by \((K T)\ell f\), then
\[
\text{Ent}_\nu(T(K T)\ell f) \leq (1 - \delta) \text{Ent}_\nu(T(K T)\ell-1 f),
\]
(7.14)
for any \(\ell \in \mathbb{N}\). Iterating this inequality we find, for any \(\ell \in \mathbb{N}\),
\[
\text{Ent}_\nu(T(K T)\ell f) \leq (1 - \delta)^\ell \text{Ent}_\nu(Tf).
\]
(7.15)
Recalling that \(P_{SW}^\ell f = (K T)^\ell f\), from \((7.15)\) we get
\[
\text{Ent}_\mu(P_{SW}^\ell f) = \text{Ent}_\nu((K T)^\ell f)
= \text{Ent}_\nu(K T(K T)^{\ell-1} f)
\leq \text{Ent}_\nu(T(K T)^{\ell-1} f)
\leq (1 - \delta)^{\ell-1} \text{Ent}_\nu(Tf)
\leq (1 - \delta)^{\ell-1} \text{Ent}_\nu(f) = (1 - \delta)^{\ell-1} \text{Ent}_\nu(f),
\]
where the first inequality follows from \((5.6)\). This shows that the discrete time entropy decay for SW on spins is asymptotically the same as the one assumed for SW on edges, and Lemma 2.4 allows us to conclude the desired mixing time bound. The same argument (with \(KT\) replaced by \(TK\)) shows that if we assume an entropy decay for spins then we obtain (asymptotically) the same entropy decay for edges, and therefore the same mixing time bound. \(\square\)
We can now provide the proof of Theorem 1.3 from the introduction.

Proof of Theorem 1.3. From the discussion at the beginning of Section 7, note that there is an admissible boundary condition in the joint space for which the edge marginal is the random-cluster measure on a square region of $\mathbb{Z}^2$ with a wired boundary condition, and the spin marginal is the monochromatic boundary condition. The result then follows from Theorem 7.3 and Lemma 7.7. □

References

[1] M. Aizenman and R. Holley. Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime. In Percolation Theory and Ergodic Theory of Infinite Particle Systems, pages 1–11. Springer, 1987.

[2] K.S. Alexander. On weak mixing in lattice models. Probability Theory and Related Fields, 110(4):441–471, 1998.

[3] C. Ané, D. Bakry, and M. Ledoux. Sur les inégalités de Sobolev logarithmiques, volume 10. Société mathématique de France, Paris, 2000.

[4] V. Beffara and H. Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. Probability Theory and Related Fields, 153:511–542, 2012.

[5] A. Blanca, P. Caputo, A. Sinclair, and E. Vigoda. Spatial mixing and non-local Markov chains. Random Structures & Algorithms, 55(3):584–614, 2019.

[6] A. Blanca, Z. Chen, and E. Vigoda. Swendsen-Wang dynamics for general graphs in the tree uniqueness region. Random Structures & Algorithms, 56(2):373–400, 2020.

[7] A. Blanca, R. Gheissari, and E. Vigoda. Random-cluster dynamics in $\mathbb{Z}^2$: rapid mixing with general boundary conditions. The Annals of Applied Probability, 30(1):418–459, 2020.

[8] A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM), pages 528–543, 2015.

[9] A. Blanca and A. Sinclair. Random-Cluster Dynamics in $\mathbb{Z}^2$. Probability Theory and Related Fields, 168:821–847, 2017.

[10] S.G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. Journal of Theoretical Probability, 19(2):289–336, 2006.

[11] C. Borgs, J.T. Chayes, A.M. Frieze, J.H. Kim, P. Tetali, E. Vigoda, and V.H. Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 218–229, 1999.

[12] C. Borgs, J. Chayes, and P. Tetali. Swendsen-Wang algorithm at the Potts transition point. Probability Theory and Related Fields, 152:509–557, 2012.

[13] P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing. arXiv preprint arXiv:2004.10574, 2020.
[14] R. Cerf and S. Louhichi. Dynamical coupling between Ising and FK percolation. *Latin American Journal of Probability and Mathematical Statistics (ALEA)*, 17:23–49, 2020.

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

[16] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statistical Science*, 23(2):151–178, 2008.

[17] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[18] M. Dyer, L.A. Goldberg, and M. Jerrum. Dobrushin conditions and systematic scan. *Combinatorics, Probability and Computing*, 17(6):761–779, 2008.

[19] M. Dyer, L.A. Goldberg, and M. Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71–107, 2009.

[20] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24:461–479, 2004.

[21] R.G. Edwards and A.D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D*, 38(6):2009–2012, 1988.

[22] C.M. Fortuin and P.W. Kasteleyn. On the random-cluster model I. Introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[23] A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures & Algorithms*, 54(1):82–147, 2019.

[24] R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean-field Swendsen-Wang dynamics. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1981–1988, 2018.

[25] V.K. Gore and M.R. Jerrum. The Swendsen-Wang process does not mix rapidly. *Journal of Statistical Physics*, 97(1-2):67–86, 1999.

[26] G.R. Grimmett. *The Random-Cluster Model*, volume 333 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006.

[27] H. Guo, K. Kara, and C. Zhang. Layerwise Systematic Scan: Deep Boltzmann Machines and Beyond. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 178–187. JMLR, 2018.

[28] T.P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 39–46, 2006.

[29] T.P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Annals of Applied Probability*, 17(3):931–952, 2007.
[30] R. Holley. Possible rates of convergence in finite range, attractive spin systems. In *Particle systems, random media and large deviations*, volume 41 of *Contemp. Math.*, pages 215–234. Amer. Math. Soc., Providence, RI, 1985.

[31] M. Huber. A bounding chain for Swendsen-Wang. *Random Structures & Algorithms*, 22(1):43–59, 2003.

[32] R. Latala and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, pages 147–168. Springer, 2000.

[33] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de probabilités XXXIII*, 120–216. Springer 1999.

[34] Y. Long, A. Nachmias, W. Ning, and Y. Peres. A power law of order 1/4 for critical mean-field Swendsen-Wang dynamics. *Memoirs of the American Mathematical Society*, 232(1092), 2011.

[35] E. Lubetzky, F. Martinelli, A. Sly, and F.L. Toninelli. Quasi-polynomial mixing of the 2D stochastic Ising model with “plus” boundary up to criticality. *Journal of the European Mathematical Society (JEMS)*, 15(2):339–3866, 2013.

[36] F. Martinelli. Dynamical analysis of low-temperature Monte Carlo cluster algorithms. *Journal of Statistical Physics*, 66(5-6):1245–1276, 1992.

[37] F. Martinelli. *Lectures on Glauber dynamics for discrete spin models*, volume 1717 of *Springer Lecture Notes in Mathematics*. Springer Verlag, 1999.

[38] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Communications in Mathematical Physics*, 161(3):447–486, 1994.

[39] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Communications in Mathematical Physics*, 161(3):458–514, 1994.

[40] F. Martinelli, E. Olivieri, and R.H. Schonmann. For 2-d lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics*, 165(1):33–47, 1994.

[41] F. Martinelli, E. Olivieri, and E. Scoppola. On the Swendsen-Wang dynamics. I. Exponential convergence to equilibrium. *Journal of Statistical Physics*, 62(1-2):117–133, 1991.

[42] F. Martinelli and F. L. Toninelli. On the mixing time of the 2D stochastic Ising model with “plus” boundary conditions at low temperature. *Communications in Mathematical Physics*, 296(1):175–213, 2010.

[43] L. Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. Séminaire de Probabilités XXXI, 136–167, 1997.

[44] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.

[45] D. Nam and A. Sly. Cutoff for the Swendsen–Wang dynamics on the lattice. *The Annals of Probability*, 47(6):3705–3761, 2019.

[46] J. Propp and D. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9:223–252, 1996.
[47] D.W. Stroock and B. Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Communications in Mathematical Physics*, 149(1):175–193, 1992.

[48] R.H. Swendsen and J.S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.

[49] M. Ullrich. Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions. *Dissertationes Mathematicae*, 502:64, 2014.

[50] B. Zegarlinski. On log-Sobolev inequalities for infinite lattice systems. *Letters in Mathematical Physics*, 20(3):173–182, 1990.