# The Swendsen-Wang Dynamics on Trees

**Antonio Blanca**
Pennsylvania State University, University Park, PA, USA

**Zongchen Chen**
Georgia Institute of Technology, Atlanta, GA, USA

**Daniel Štefankovič**
University of Rochester, NY, USA

**Eric Vigoda**
Georgia Institute of Technology, Atlanta, GA, USA

## Abstract

The Swendsen-Wang algorithm is a sophisticated, widely-used Markov chain for sampling from the Gibbs distribution for the ferromagnetic Ising and Potts models. This chain has proved difficult to analyze, due in part to the global nature of its updates. We present optimal bounds on the convergence rate of the Swendsen-Wang algorithm for the complete $d$-ary tree. Our bounds extend to the non-uniqueness region and apply to all boundary conditions. We show that the spatial mixing conditions known as *Variance Mixing* and *Entropy Mixing*, introduced in the study of local Markov chains by Martinelli et al. (2003), imply $\Omega(1)$ spectral gap and $O(\log n)$ mixing time, respectively, for the Swendsen-Wang dynamics on the $d$-ary tree. We also show that these bounds are asymptotically optimal. As a consequence, we establish $\Theta(\log n)$ mixing for the Swendsen-Wang dynamics for *all* boundary conditions throughout the tree uniqueness region; in fact, our bounds hold beyond the uniqueness threshold for the Ising model, and for the $q$-state Potts model when $q$ is small with respect to $d$. Our proofs feature a novel spectral view of the Variance Mixing condition inspired by several recent rapid mixing results on high-dimensional expanders and utilize recent work on block factorization of entropy under spatial mixing conditions.

**2012 ACM Subject Classification** Theory of computation → Random walks and Markov chains; Mathematics of computing → Markov processes; Theory of computation → Design and analysis of algorithms

**Keywords and phrases** Markov Chains, mixing times, Ising and Potts models, Swendsen-Wang dynamics, trees

**Digital Object Identifier** 10.4230/LIPIcs.APPROX/RANDOM.2021.43

**Category** RANDOM

**Related Version** *Full Version*: https://arxiv.org/abs/2007.08068

**Funding** *Antonio Blanca*: Research supported in part by NSF grant CCF-1850443.
*Zongchen Chen*: Research supported in part by NSF grant CCF-2007022.
*Daniel Štefankovič*: Research supported in part by NSF grant CCF-2007287.
*Eric Vigoda*: Research supported in part by NSF grant CCF-2007022.

## 1 Introduction

Spin systems are idealized models of a physical system in equilibrium which are utilized in statistical physics to study phase transitions. A phase transition occurs when there is a dramatic change in the macroscopic properties of the system resulting from a small (infinitesimal in the limit) change in one of the parameters defining the spin system. The macroscopic properties of the system manifest with the persistence (or lack thereof) of long-


Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021).
Editors: Mary Wootters and Laura Sanità; Article No. 43; pp. 43:1–43:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

range influences. There is a well-established mathematical theory connecting the absence of these influences to the fast convergence of Markov chains. In this paper, we study this connection on the regular tree, known as the *Bethe lattice* in statistical physics [7, 27].

The most well-studied example of a spin system is the *ferromagnetic q-state Potts model*, which contains the *Ising model* ($q = 2$) as a special case. The Potts model is especially important as fascinating phase transitions (first-order vs. second-order) are now understood rigorously in various contexts [5, 21, 20, 17, 18].

Given a graph $G = (V, E)$, configurations of the Potts model are assignments of spins $[q] = \{1, 2, \dots, q\}$ to the vertices of $G$. The parameter $\beta > 0$ (corresponding to the inverse of the temperature of the system) controls the strength of nearest-neighbor interactions, and the probability of a configuration $\sigma \in [q]^V$ in the *Gibbs distribution* is such that

$$\mu(\sigma) = \mu_G(\sigma) = \frac{e^{-\beta|D(\sigma)|}}{Z}, \tag{1}$$

where $D(\sigma) = \{\{v, w\} \in E : \sigma(v) \neq \sigma(w)\}$ denotes the set of bi-chromatic edges in $\sigma$, and $Z$ is the normalizing constant known as the *partition function.*

The *Glauber dynamics* is the simplest example of a Markov chain for sampling from the Gibbs distribution; it updates the spin at a randomly chosen vertex in each step. In many settings, as we detail below, the Glauber dynamics converges exponentially slow at low temperatures (large $\beta$) due to the local nature of its transitions and the long-range correlations in the Gibbs distribution. Of particular interest are thus "global" Markov chains such as the *Swendsen-Wang (SW) dynamics* [49, 23], which update a large fraction of the configuration in each step, thus potentially overcoming the obstacles that hinder the performance of the Glauber dynamics, and with steps that can be efficiently parallelized [4].

The SW dynamics utilizes a close connection between the Potts model and an alternative representation known as the *random-cluster model.* The random-cluster model is defined on subsets of edges and is not a spin system as the weight of a configuration depends on the global connectivity properties of the corresponding subgraph. The transitions of the SW dynamics take a spin configuration, transform it to a "joint" spin-edge configuration, perform a step in the joint space, and then map back to a Potts configuration. Formally, from a Potts configuration $\sigma_t \in [q]^V$, a transition $\sigma_t \to \sigma_{t+1}$ is defined as follows:

1. Let $M_t = M(\sigma_t) = E \setminus D(\sigma_t)$ denote the set of monochromatic edges in $\sigma_t$.
2. Independently for each edge $e = \{v, w\} \in M_t$, keep $e$ with probability $p = 1 - \exp(-\beta)$ and remove $e$ with probability $1 - p$. Let $A_t \subseteq M_t$ denote the resulting subset.
3. In the subgraph $(V, A_t)$, independently for each connected component $C$ (including isolated vertices), choose a spin $s_C$ uniformly at random from $[q]$ and assign to each vertex in $C$ the spin $s_C$. This spin assignment defines $\sigma_{t+1}$.

There are two standard measures of the convergence rate of a Markov chain. The *mixing time* is the number of steps to get within total variation distance $\leq 1/4$ of its stationary distribution from the worst starting state. The *relaxation time* is the inverse of the spectral gap of the transition matrix of the chain and measures the speed of convergence from a "warm start". For approximate counting algorithms the relaxation time is quite useful as it corresponds to the "resample" time [30, 37, 35, 34]; see Section 2 for precise definitions and how these two notions relate to each other.

There has been great progress in formally connecting phase transitions with the convergence rate of the Glauber dynamics. Notably, for the $d$-dimensional integer lattice $\mathbb{Z}^d$, a series of works established that a spatial mixing property known as *strong spatial mixing (SSM)* implies $O(n \log n)$ mixing time of the Glauber dynamics [39, 15, 22]. Roughly speaking, SSM says that correlations decay exponentially fast with the distance and is also known to imply optimal mixing and relaxation times of the SW dynamics on $\mathbb{Z}^d$ [9, 8]. These techniques utilizing SSM are particular to the lattice and do not extend to *non-amenable* graphs (i.e., those whose boundary and volume are of the same order). The $d$-ary complete tree, which is the focus of this paper, is the prime example of a non-amenable graph.

On the regular $d$-ary tree, there are two fundamental phase transitions: the *uniqueness* threshold $\beta_u$ and the *reconstruction* threshold $\beta_r$. The smaller of these thresholds $\beta_u$ corresponds to the uniqueness/non-uniqueness phase transition of the Gibbs measure on the infinite $d$-ary tree, and captures whether the worst-case boundary configuration (i.e., a fixed configuration on the leaves of a finite tree) has an effect or not on the spin at the root (in the limit as the height of the tree grows). The second threshold $\beta_r$ is the reconstruction/non-reconstruction phase transition, marking the divide on whether or not a *random* boundary condition (in expectation) affects the spin of the root.

There is a large body of work on the interplay between these phase transitions and the speed of convergence of the Glauber dynamics on the complete $d$-ary tree [41, 40, 6], and more generally on bounded degree graphs [44, 28, 13]. Our main contributions in this paper concern instead the speed of convergence of the SW dynamics on trees, how it is affected by these phase transitions, and the effects of the boundary condition.

Martinelli, Sinclair, and Weitz [40, 41] introduced a pair of spatial mixing (decay of correlation) conditions called *Variance Mixing (VM)* and *Entropy Mixing (EM)* which capture the exponential decay of point-to-set correlations. More formally, the VM and EM conditions hold when there exist constants $\ell > 0$ and $\varepsilon = \varepsilon(\ell)$ such that, for every vertex $v \in T$, the influence of the spin at $v$ on the spins of the vertices at distance $\geq \ell$ from $v$ in the subtree $T_v$ rooted at $v$ decays by a factor of at least $\varepsilon$. For the case of VM, this decay of influence is captured in terms of the variance of any function $g$ that depends only on the spins of the vertices in $T_v$ at distance $\geq \ell$ from $v$; specifically, when conditioned on the spin at $v$, the conditional variance of $g$ is (on average) a factor $\varepsilon$ smaller then the unconditional variance; see Definition 7 in Section 3 for the formal definition. EM is defined analogously, with variance replaced by entropy; see Definition 15.

It was established in [40, 41] that VM and EM imply optimal bounds on the convergence rate of the Glauber dynamics on trees. We obtain optimal bounds for the speed of convergence of the SW dynamics under the same VM and EM spatial mixing conditions.

▶ **Theorem 1.** *For all $q \geq 2$ and $d \geq 3$, for the $q$-state ferromagnetic Ising/Potts model on an $n$-vertex complete $d$-ary tree, Variance Mixing implies that the relaxation time of the Swendsen-Wang dynamics is $\Theta(1)$.*

▶ **Theorem 2.** *For all $q \geq 2$ and $d \geq 3$, for the $q$-state ferromagnetic Ising/Potts model on an $n$-vertex complete $d$-ary tree, Entropy Mixing implies that the mixing time of the Swendsen-Wang dynamics is $O(\log n)$.*

The VM condition is strictly weaker (i.e., easier to satisfy) than the EM condition, but, at the moment, EM is known to hold in the same parameter regimes as VM. The relaxation time bound in Theorem 1 is weaker than the mixing time bound in Theorem 2. We also show that the mixing time in Theorem 2 is asymptotically the best possible.

▶ **Theorem 3.** *For all $q \geq 2$, $d \geq 3$ and any $\beta > 0$, the mixing time of the SW dynamics on an $n$-vertex complete $d$-ary tree is $\Omega(\log n)$ for any boundary condition.*

We remark that the mixing time lower bound in Theorem 3 applies to all inverse temperatures $\beta$ and all boundary conditions.

The VM and EM conditions are properties of the Gibbs distribution induced by *a specific* boundary condition on the leaves of the tree; this contrasts with other standard notions of decay of correlations such as SSM on $\mathbb{Z}^d$. This makes these conditions quite suitable for understanding the speed of convergence of Markov chains under different boundary conditions. For instance, [40, 41] established VM and EM for all boundary conditions provided $\beta < \max\{\beta_u, \frac{1}{2}\ln(\frac{\sqrt{d}+1}{\sqrt{d}-1})\}$ and for the monochromatic (e.g., all-red) boundary condition for all $\beta$. Consequently, we obtain the following results.

▶ **Theorem 4.** *For all $q \geq 2$ and $d \geq 3$, for the $q$-state ferromagnetic Ising/Potts model on an $n$-vertex complete $d$-ary tree, the relaxation time of the Swendsen-Wang dynamics is $\Theta(1)$ and its mixing time is $\Theta(\log n)$ in the following cases:*
1. *the boundary condition is arbitrary and $\beta < \max\left\{\beta_u, \frac{1}{2}\ln\left(\frac{\sqrt{d}+1}{\sqrt{d}-1}\right)\right\}$;*
2. *the boundary condition is monochromatic and $\beta$ is arbitrary.*

Part (i) of this theorem provides optimal mixing and relaxation times bounds for the SW dynamics under arbitrary boundaries throughout the uniqueness region $\beta < \beta_u$. In fact, $\beta_u < \frac{1}{2}\ln(\frac{\sqrt{d}+1}{\sqrt{d}-1})$ when $q \leq 2(\sqrt{d}+1)$ and thus our bound extends to the non-uniqueness region for many combinations of $d$ and $q$. We note that while the value of the uniqueness threshold $\beta_u$ is known, it does not have a closed form (see [31, 10]). In contrast, the reconstruction threshold $\beta_r$ is not known for the Potts model [47, 43], but one would expect that part (i) holds for all $\beta < \beta_r$; analogous results are known for the Glauber dynamics for other spin systems where more precise bounds on the reconstruction threshold have been established [6, 45, 48].

Previously, only a poly($n$) bound was known for the mixing time of the SW dynamics for arbitrary boundary conditions [50, 6]. This poly($n$) bound holds for every $\beta$, but the degree of the polynomial bounding the mixing time is quite large (grows with $\beta$); our bound in part (i) is thus a substantial improvement.

In regards to part (ii) of the theorem, we note that our bound holds for all $\beta$, including the whole low-temperature region. The only other case where tight bounds for the SW dynamics are known for the full low-temperature regime is on the geometrically simpler complete graph [25, 12].

Previous (direct) analysis of the speed of convergence of the SW dynamics on trees focused exclusively on the special case of the *free* boundary condition [33, 16], where the dynamics is much simpler as the corresponding random-cluster model is trivial (reduces to independent bond percolation); this was used by Huber [33] to establish $O(\log n)$ mixing time of the SW dynamics for all $\beta$ for the special case of the free boundary condition.

We comment briefly on our proof methods next; a more detailed exposition of our approach is provided later in this introduction. The results in [40, 41] use the VM and EM condition to deduce optimal bounds for the relaxation and mixing times of the Glauber dynamics; specifically, they analyze its spectral gap and log-Sobolev constant. Their methods do not extend to the SW dynamics. It can be checked, for example, that the log-Sobolev constant for the SW dynamics is $\Theta(n^{-1})$, and thus the best possible mixing time one could hope to obtain with such an approach would be $O(n \log n)$. For Theorem 2, we utilize instead new tools introduced by Caputo and Parisi [14] to establish a (block) factorization of entropy. This factorization allows to get a handle on the *modified* log-Sobolev constant for the SW dynamics. For Theorem 1, the main novelty in our approach is a new spectral interpretation of the VM condition that facilitates a factorization of variance, similar to the factorization of entropy from [14]. Lastly, the lower bound from Theorem 3 is obtained by adapting the framework of Hayes and Sinclair [32] to the SW setting using recent ideas from [8].

Finally, we mention that part (ii) of Theorem 4 has interesting implications related to the speed of convergence of random-cluster model Markov chains on trees under the wired boundary condition. That is, all the leaves are connected through external or "artificial" wirings. The case of the wired boundary condition is the most studied version of the random-cluster model on trees (see, e.g., [31, 36]) since, as mentioned earlier, the model is trivial under the free boundary. The random-cluster model, which is parameterized by $p \in (0,1)$ and $q > 0$ (see [24, 1] for its definition), is intimately connected to the ferromagnetic $q$-sate Potts model when $q \geq 2$ is an integer and $p = 1 - \exp(-\beta)$. In particular, there is a variant of SW dynamics for the random-cluster model (by observing the edge configuration after the second step of the chain).

Another standard Markov chain for the random-cluster model is the heat-bath (edge) dynamics, which is the analog of the Glauber dynamics on spins for random-cluster configurations. Our results for the random-cluster dynamics are the following.

▶ **Theorem 5.** *For all integer $q \geq 2$, $p \in (0,1)$, and $d \geq 3$, for the random-cluster model on an $n$-vertex complete $d$-ary tree with* wired boundary condition*, the mixing time of the Swendsen-Wang dynamics is $O(\log n)$. In addition, the mixing time of the heat-bath edge dynamics for the random-cluster model is $O(n \log n)$.*

To prove these results, we use a factorization of entropy in the joint spin-edge space, as introduced in [8]; they cannot be deduced from the mixing time bounds for the Glauber dynamics for the Potts model in [40, 41].

Our final result shows that while random-cluster dynamics mix quickly under the wired boundary condition, there are random-cluster boundary conditions that cause an exponential slowdown for both the SW dynamics and the heat-bath edge dynamics for the random-cluster model.

▶ **Theorem 6.** *For all $q \geq 2$, all $d \geq 3$, consider the random-cluster model on an $n$-vertex complete $d$-ary tree. Then, there exists $p \in (0,1)$ and a random-cluster boundary condition such that the mixing times of the Swendsen-Wang dynamics and of the heat-bath edge dynamics is $\exp(\Omega(\sqrt{n}))$.*

We prove this result extending ideas from [11]. In particular, we prove a general theorem that allows us to transfer slow mixing results for the edge dynamics on other graphs to the tree, for a carefully constructed tree boundary condition and a suitable $p$. To proof this results we use the random-cluster boundary condition to embed an arbitrary graph $G$ on the tree; a set with bad conductance for the chain on $G$ is then lifted to the tree. Theorem 6 then follows from any of the known slow mixing results for the edge dynamics [29, 26, 50].

**Our techniques.** Our first technical contribution is a reinterpretation and generalization of the VM condition as a bound on the second eigenvalue of a certain stochastic matrix which we denote by $P^{\uparrow}P^{\downarrow}$. The matrices $P^{\uparrow}$ and $P^{\downarrow}$ are distributional matrices corresponding to the distribution at a vertex $v$ given the spin configuration of the set $S_v$ of all its descendants at distance at least $\ell$ and vice versa. These matrices are inspired by the recent results in [2, 3] utilizing high-dimensional expanders; see Section 3 for their precise definitions.

Our new spectral interpretation of the VM condition allows us to factorize it and obtain an equivalent global variant we call *Parallel Variance Mixing (PVM)*. While the VM condition signifies the exponential decay with distance of the correlations between a vertex $v$ and the set $S_v$ (and is well-suited for the analysis of local Markov chains), the PVM condition captures instead the decay rate of set-to-set correlations between the set of all the vertices at a fixed level of the tree and the set of all their descendants at distance at least $\ell$. The PVM condition facilitates the analysis of a block dynamics with a constant number of blocks each of linear volume. We call this variant of block dynamics the *tiled block dynamics* as each block consists of a maximal number of non-intersecting subtrees of constant size (i.e., a *tiling*); see Figure 1. We use the PVM condition to show that the spectral gap of the tiled block dynamics is $\Omega(1)$, and a generic comparison between the block dynamics and the SW dynamics yields Theorem 1.

Our proof of Theorem 2 follows a similar strategy. We first obtain a global variant of the EM condition, analogous to the PVM condition but for entropy. For this, we use a recent result of Caputo and Parisi [14]. From this global variant of the EM condition we deduce a factorization of entropy into the even and odd subsets of vertices. (The parity of a vertex is that of its distance to the leaves of the tree.) The even-odd factorization of entropy was recently shown in [8] to imply $O(\log n)$ mixing of the SW on general biparte graphs.

**Paper organization.** The rest of the paper is organized as follows. Section 2 contains some standard definitions and facts we use in our proofs. In Sections 3 and 4 we prove Theorems 1 and 2, respectively. Our general comparison result between the SW dynamics and the block dynamics, our results for the random-cluster model dynamics, and our lower bound for the SW dynammics (Theorem 3) are proved in the full version of this paper [1].

## 2 Preliminaries

We introduce some notations and facts that are used in the remainder of the paper.

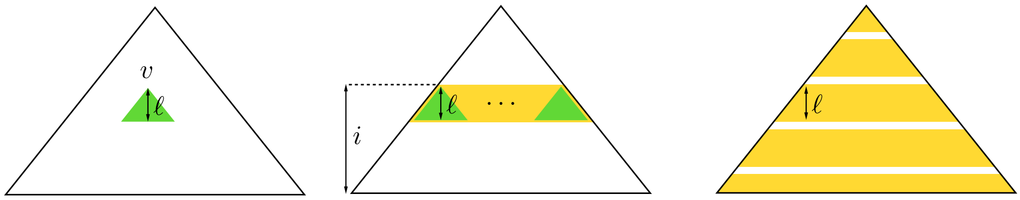
**The Potts model on the $d$-ary tree.** For $d \geq 2$, let $\mathbb{T}^d = (\mathbb{V}, \mathbb{E})$ denote the rooted infinite $d$-ary tree in which every vertex (including the root) has exactly $d$ children. We consider the complete finite subtree of $\mathbb{T}^d$ of height $h$, which we denote by $T = T_h^d = (V(T), E(T))$. We use $\partial T$ to denote the external boundary of $T$; i.e., the set of vertices in $\mathbb{V} \setminus V(T)$ incident to the leaves of $T$. We identify subgraphs of $T$ with their vertex sets. In particular, for $A \subseteq V(T)$ we use $E(A)$ for the edges with both endpoints in $A$, $\partial A$ for the external boundary of $A$ (i.e., the vertices in $(T \cup \partial T) \setminus A$ adjacent to $A$), and, with a slight abuse of notation, we write $A$ also for the induced subgraph $(A, E(A))$. When clear from context, we simply use $T$ for the vertex set $V(T)$.

A configuration of the Potts model is an assignment of spins $[q] = \{1, \dots, q\}$ to the vertices of the graph. For a fixed spin configuration $\tau$ on the infinite tree $\mathbb{T}^d$, we use $\Omega^\tau = [q]^{T \cup \partial T}$ to denote the set of configurations of $T$ that agree with $\tau$ on $\partial T$. Hence, $\tau$ specifies a boundary condition for $T$. More generally, for any $A \subseteq T$ and any $\eta \in \Omega^\tau$, let $\Omega_A^\eta \subseteq \Omega^\tau$ denote the set of configurations of $T$ that agree with $\eta$ on $(T \cup \partial T) \setminus A$. We use $\mu_A^\eta$ to denote the Gibbs distribution over $\Omega_A^\eta$, so for $\sigma \in \Omega_A^\eta$ we have

$$\mu_A^\eta(\sigma) := \frac{1}{Z} \exp\Big( -\beta \sum\nolimits_{\{u,v\} \in E(A \cup \partial A)} \mathbb{1}(\sigma_u \neq \sigma_v) \Big),$$

where $Z$ is a normalizing constant (or partition function). For $\sigma \notin \Omega_A^\eta$, we set $\mu_A^\eta(\sigma) = 0$.

**The tiled block dynamics.** Let $\mathcal{U} = \{U_1, ..., U_r\}$ be a collection of subsets (or blocks) such that $T = \bigcup_i U_i$. The (heat-bath) block dynamics with blocks $\mathcal{U}$ is a standard Markov chain for the Gibbs distribution $\mu_T^\tau$. If the configuration at time $t$ is $\sigma_t$, the next configuration $\sigma_{t+1}$ is generated as follows:




**(a)** $B(v,\ell)$. **(b)** $B_i^\ell$. **(c)** $T_j^\ell$.

**Figure 1** An illustration of the sets $B(v,\ell)$, $B_i^\ell$, and $T_j^\ell$, where $\ell$ represents the number of levels.

1. Pick an integer $j \in \{1,2,\dots,r\}$ uniformly at random;
2. Draw a sample $\sigma_{t+1}$ from the conditional Gibbs distribution $\mu_{U_j}^{\sigma_t}$; that is, update the configuration in $U_j$ with a new configuration distributed according to the conditional measure in $U_j$ given the configuration of $\sigma_t$ on $(T \cup \partial T) \setminus U_j$ and the boundary condition $\tau$.

We consider a special choice of blocks, where each block is a disjoint union of small subtrees of constant height forming a tiling structure. For $0 \le i \le h+1$, let $L_i$ denote the set of vertices of $T$ that are of distance exactly $i$ from the boundary $\partial T$; in particular, $L_0 = \emptyset$ and $L_{h+1}$ contains only the root of $T$. (It will be helpful to define $L_i = \emptyset$ for $i < 0$ or $i > h+1$.) Let $F_i = \cup_{j \le i} L_j$ be the set of vertices at distance at most $i$ from $\partial T$; then $F_0 = \emptyset$ and $F_{h+1} = T$. We further define $F_i = \emptyset$ for $i < 0$ and $F_i = T$ for $i > h+1$. For each $i \in \mathbb{N}^+$ let

$$B_i^\ell = F_i \backslash F_{i-\ell} = \bigcup_{i-\ell < j \le i} L_j. \tag{2}$$

In words, $B_i^\ell$ is the collection of all the subtrees of $T$ of height $\ell - 1$ with roots at distance exactly $i$ from $\partial T$; see Figure 1(b). Finally, for each $1 \le j \le \ell+1$, we define

$$T_j^\ell = \bigcup_{0 \le k \le \frac{h+\ell-j}{\ell+1}} B_{j+k(\ell+1)}^\ell. \tag{3}$$

The set $T_j^\ell$ contains all the subtrees of $T$ whose roots are at distance $j + k(\ell+1)$ from $\partial T$ for some non-negative integer $k$; the height of each subtree (except the top and bottom ones) is $\ell - 1$. Also notice that all the subtrees in $T_j^\ell$ are at (graph) distance at least 2 from each other, and thus they create a tiling pattern over $T$. Therefore, we call the block dynamics with blocks $\mathcal{U} = \{T_1^\ell, \dots, T_{\ell+1}^\ell\}$ the *tiled block dynamics*; see Figure 1(c). The transition matrix of the tiled block dynamics is denoted by $P_{\text{TB}}$.

**Mixing and relaxation times.** Let $P$ be the transition matrix of an ergodic Markov chain over a finite set $\Phi$ with stationary distribution $\nu$. We use $P^t(X_0,\cdot)$ to denote the distribution of the chain after $t$ steps starting from $X_0 \in \Phi$. The mixing time of $P$ is defined as $\tau_{\text{mix}}(P) = \max_{X_0 \in \Phi} \min\{t \ge 0 : \|P^t(X_0,\cdot) - \nu\|_{\text{TV}} \le 1/4\}$, where $\|\cdot\|_{\text{TV}}$ denotes total variation distance.

When $P$ is reversible, its spectrum is real and we let $1 = \lambda_1 > \lambda_2 \ge \dots \ge \lambda_{|\Phi|} \ge -1$ denote its eigenvalues ($1 > \lambda_2$ when $P$ is irreducible). The *absolute spectral gap* of $P$ is defined by $\mathsf{gap}(P) = 1 - \lambda^*$, where $\lambda^* = \max\{|\lambda_2|, |\lambda_{|\Phi|}|\}$. If $P$ is ergodic (i.e., irreducible and aperiodic), then $\mathsf{gap}(P) > 0$, and it is a standard fact that if $\nu_{\min} = \min_{x\in\Phi} \nu(x)$, then

$$\left(\mathsf{gap}(P)^{-1} - 1\right)\log 2 \le \tau_{\text{mix}}(P) \le \mathsf{gap}(P)^{-1} \log\left(4\nu_{\min}^{-1}\right); \tag{4}$$

see [38]. The *relaxation time* of the chain is defined as $\mathsf{gap}(P)^{-1}$.

**Analytic tools.** We review next some useful tools from functional analysis; we refer the reader to [42, 46] for more extensive background. We can endow $\mathbb{R}^\Phi$ with the inner product $\langle f,g\rangle_\nu = \sum_{x\in\Phi} f(x)g(x)\nu(x)$ for two functions $f,g:\Phi\to\mathbb{R}$. The resulting Hilbert space is denoted by $L_2(\nu) = (\mathbb{R}^\Phi, \langle\cdot,\cdot\rangle_\nu)$ and $P$ defines an operator from $L_2(\nu)$ to $L_2(\nu)$.

Let $\mathbf{1}:\Phi\to\mathbb{R}$ be the constant "all 1" function (i.e., $\mathbf{1}(x)=1\ \forall x\in\Phi$) and let $I$ denote the identity mapping over all functions (i.e., $If=f$ for all $f:\Phi\to\mathbb{R}$). We then define:

$$\mathbb{E}_\nu(f) = \sum_{x\in\Phi} f(x)\nu(x) = \langle f,\mathbf{1}\rangle_\nu\,, \text{ and}$$
$$\mathrm{Var}_\nu(f) = \mathbb{E}_\nu(f^2) - \mathbb{E}_\nu(f)^2 = \langle f,(I-\mathbf{1}\nu)f\rangle_\nu$$

as the expectation and variance of the function $f$ with respect to (w.r.t.) the measure $\nu$. Likewise, for a function $f:\Omega\to\mathbb{R}_{\geq 0}$ we define the entropy of $f$ with respect to $\nu$ as $\mathrm{Ent}_\nu(f) = \mathbb{E}_\nu\left[f\log\left(\frac{f}{\mathbb{E}_\nu(f)}\right)\right]$.

Often, we will consider $\nu$ to be the conditional Gibbs distribution $\mu_A^\eta$ for some $A\subseteq T$ and $\eta\in\Omega$. In those cases, to simplify the notation, we shall write $\mathbb{E}_A^\eta(f)$ for $\mathbb{E}_{\mu_A^\eta}(f)$, $\mathrm{Var}_A^\eta(f)$ for $\mathrm{Var}_{\mu_A^\eta}(f)$, and $\mathrm{Ent}_A^\eta(f)$ for $\mathrm{Ent}_{\mu_A^\eta}(f)$.

The *Dirichlet form* of a reversible Markov chain with transition matrix $P$ is defined as

$$\mathcal{E}_P(f,f) = \langle f,(I-P)f\rangle_\nu = \frac{1}{2}\sum_{x,y\in\Phi}\nu(x)P(x,y)(f(x)-f(y))^2, \tag{5}$$

for any $f:\Phi\to\mathbb{R}$. We say $P$ is *positive semidefinite* if $\langle f,Pf\rangle_\nu\geq 0$ for all functions $f:\Phi\to\mathbb{R}$. In this case $P$ has only nonnegative eigenvalues. If $P$ is positive semidefinite, then the absolute spectral gap of $P$ satisfies

$$\mathsf{gap}(P) = 1-\lambda_2 = \inf_{\substack{f:\Phi\to\mathbb{R}\\ \mathrm{Var}_\nu(f)\neq 0}} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\nu(f)}. \tag{6}$$

## 3 Variance Mixing implies fast mixing: Proof of Theorem 1

We start with the formal definition of the *Variance Mixing (VM)* condition introduced by Martinelli, Sinclair and Weitz [40]. Throughout this section, we consider the Potts model on the $n$-vertex $d$-ary complete tree $T=T_h^d$ with a fixed boundary condition $\tau$; hence, for ease of notation we set $\mu := \mu_T^\tau$ and $\Omega := \Omega^\tau$.

For $v\in T$, let $T_v$ denote the subtree of $T$ rooted at $v$. For boundary condition $\eta\in\Omega$ and a function $g:\Omega_{T_v}^\eta\to\mathbb{R}$, we define the function $g_v:[q]\to\mathbb{R}$ as the conditional expectation

$$g_v(a) = \mathbb{E}_{T_v}^\eta[g\mid\sigma_v=a] = \sum_{\sigma\in\Omega_{T_v}^\eta:\sigma_v=a}\mu_{T_v}^\eta(\sigma\mid\sigma_v=a)g(\sigma). \tag{7}$$

In words, $g_v(a)$ is the conditional expectation of the function $g$ under the distribution $\mu_{T_v}^\eta$ given that the root of $T_v$ (i.e, the vertex $v$) is set to spin $a\in[q]$. We also consider the expectation and variance of $g_v$ w.r.t. the projection of $\mu_{T_v}^\eta$ on $v$. In particular,

$$\mathbb{E}_{T_v}^\eta[g_v] = \sum\nolimits_{a\in[q]}\mu_{T_v}^\eta(\sigma_v=a)g_v(a) = \mathbb{E}_{T_v}^\eta[g], \text{ and}$$
$$\mathrm{Var}_{T_v}^\eta[g_v] = \mathbb{E}_{T_v}^\eta[g_v^2] - \mathbb{E}_{T_v}^\eta[g_v]^2.$$

For an integer $\ell\geq 1$, we define $B(v,\ell)$ as the set of vertices of $T_v$ that are at distance less than $\ell$ from $v$; see Figure 1(a). We say that the function $g:\Omega_{T_v}^\eta\to\mathbb{R}$ is independent of the configuration on $B(v,\ell)$ if for all $\sigma,\sigma'\in\Omega_{T_v}^\eta$ such that $\sigma(B(v,\ell))\neq\sigma'(B(v,\ell))$ and $\sigma(T_v\setminus B(v,\ell)) = \sigma'(T_v\setminus B(v,\ell))$, we have $g(\sigma)=g(\sigma')$. We can now define VM.

▶ **Definition 7** (Variance Mixing (VM)). *The Gibbs distribution $\mu = \mu_T^\tau$ satisfies* $\mathrm{VM}(\ell, \varepsilon)$ *if for every $v \in T$, every $\eta \in \Omega$, and every function $g : \Omega_{T_v}^\eta \to \mathbb{R}$ that is independent of the configuration on $B(v, \ell)$, we have $\mathrm{Var}_{T_v}^\eta(g_v) \le \varepsilon \cdot \mathrm{Var}_{T_v}^\eta(g)$. We say that the VM condition holds if there exist constants $\ell$ and $\varepsilon = \varepsilon(\ell)$ such that* $\mathrm{VM}(\ell, \varepsilon)$ *holds.*

The VM condition is a spatial mixing property that captures the rate of decay of correlations, given by $\varepsilon = \varepsilon(\ell)$, with the distance $\ell$ between $v \in T$ and the set $T_v \setminus B(v, \ell)$. To see this, note that, roughly speaking, $\mathrm{Var}_{T_v}^\eta(g_v)$ is small when $g_v(a) = \mathbb{E}_{T_v}^\eta[g \mid \sigma_v = a]$ is close to $g_v(b) = \mathbb{E}_{T_v}^\eta[g \mid \sigma_v = b]$ for every $a \neq b$. Since $g$ is independent of the configuration on $B(v, \ell)$, this can only happen if the spin at $v$, which is at distance $\ell$ from $T_v \setminus B(v, \ell)$, has only a small influence on the projections of the conditional measures $\mu_{T_v}^\eta(\cdot \mid \sigma_v = a)$, $\mu_{T_v}^\eta(\cdot \mid \sigma_v = b)$ to $T_v \setminus B(v, \ell)$.

It was established in [40, 41] that VM implies optimal mixing of the Glauber dynamics; this was done by analyzing a block dynamics that updates one random block $B(v, \ell)$ in each step. This block dynamics behaves similarly to the Glauber dynamics since all blocks are of constant size, and there are a linear number of them; see [40, 41] for further details. Our goal here is to establish optimal mixing of global Markov chains, and thus we require a different spatial mixing condition that captures decay of correlations in a more global manner. For this, we introduce the notion of *Parallel Variance Mixing (PVM)*. Recall that for $0 \le i \le h+1$, $L_i$ is the set all vertices at distance exactly $i$ from the boundary $\partial T$, $F_i = \cup_{j \le i} L_j$, and $B_i^\ell = F_i \backslash F_{i-\ell}$; see Figures 1(b) and 1(c).

For $1 \le i \le h+1$, $\eta \in \Omega$ and $g : \Omega_{F_i}^\eta \to \mathbb{R}$, consider the function $g_{L_i} : [q]^{L_i} \to \mathbb{R}$ given by

$$g_{L_i}(\xi) = \mathbb{E}_{F_i}^\eta[g \mid \sigma_{L_i} = \xi] = \sum_{\sigma \in \Omega_{F_i}^\eta : \sigma_{L_i} = \xi} \mu_{F_i}^\eta(\sigma \mid \sigma_{L_i} = \xi) g(\sigma),$$

for $\xi \in [q]^{L_i}$. That is, $g_{L_i}(\xi)$ is the conditional expectation of function $g$ under the distribution $\mu_{T_v}^\eta$ conditioned on the configuration of the level $L_i$ being $\xi$. Thus, we may consider the expectation and variance of $g_{L_i}$ w.r.t. the projection of $\mu_{T_v}^\eta$ to $L_i$; namely, $\mathbb{E}_{F_i}^\eta[g_{L_i}] = \mathbb{E}_{F_i}^\eta[g]$ and $\mathrm{Var}_{F_i}^\eta[g_{L_i}] = \mathbb{E}_{F_i}^\eta[g_{L_i}^2] - \mathbb{E}_{F_i}^\eta[g_{L_i}]^2$. The PVM condition is defined as follows.

▶ **Definition 8** (Parallel Variance Mixing (PVM)). *The Gibbs distribution $\mu = \mu_T^\tau$ satisfies* $\mathrm{PVM}(\ell, \varepsilon)$ *if for every $1 \le i \le h+1$, every $\eta \in \Omega$, and every function $g : \Omega_{F_i}^\eta \to \mathbb{R}$ that is independent of the configuration on $B_i^\ell$, we have $\mathrm{Var}_{F_i}^\eta(g_{L_i}) \le \varepsilon \cdot \mathrm{Var}_{F_i}^\eta(g)$. The PVM condition holds if there exist constants $\ell$ and $\varepsilon = \varepsilon(\ell)$ such that* $\mathrm{PVM}(\ell, \varepsilon)$ *holds.*

PVM is a natural global variant of VM since $F_i = \bigcup_{v \in L_i} T_v$ and $B_i^\ell = \bigcup_{v \in L_i} B(v, \ell)$. We can show that the two properties are actually equivalent.

▶ **Theorem 9.** *For every $\ell \in \mathbb{N}^+$ and $\varepsilon \in (0, 1)$, the Gibbs distribution $\mu$ satisfies* $\mathrm{VM}(\ell, \varepsilon)$ *if and only if $\mu$ satisfies* $\mathrm{PVM}(\ell, \varepsilon)$.

In order to show the equivalence between VM and PVM, we introduce a more general spatial mixing condition which we call *General Variance Mixing (GVM)*. We define GVM for general product distributions (see Definition 12) and reinterpret VM and PVM as special cases of this condition. This alternative view of VM and PVM in terms of GVM is quite useful since we can recast the GVM condition as a bound on the spectral gap of a certain Markov chain; this is one key insight in the proof of Theorem 5 and is discussed in detail in Section 3.1.

Now, while VM implies optimal mixing of the Glauber dynamics, we can show that PVM implies a constant bound on the spectral gap of the tiled block dynamics. Recall that this is the heat-bath block dynamics with block collection $\mathcal{U} = \{T_1^\ell, \dots, T_{\ell+1}^\ell\}$ defined in Section 2.

▶ **Theorem 10.** *If there exist $\ell \in \mathbb{N}^+$ and $\delta \in (0,1)$ such that $\mu = \mu_T^\tau$ satisfies* $\mathrm{PVM}(\ell,\varepsilon)$ *for $\varepsilon = \frac{1-\delta}{2(\ell+1)}$, then the relaxation time of the tiled block dynamics is at most $2(\ell+1)/\delta$.*

To prove Theorem 10, we adapt the methods from [40, 41] to our global setting. Our result for the spectral gap of the SW dynamics (Theorem 1) is then obtained through comparison with the tiled block dynamics. We prove the following comparison result between the SW dynamics and a large class of block dynamics, which could be of independent interest.

▶ **Theorem 11.** *Let $\mathcal{D} = \{D_1,\dots,D_m\}$ be such that $D_i \subseteq T$ and $\cup_{i=1}^m D_i = T$. Suppose that each block $D_k$ is such that $D_k = \cup_{j=1}^{\ell_k} D_{kj}$ where $\mathrm{dist}(D_{kj}, D_{kj'}) \ge 2$ for every $j \neq j'$ and let $\mathrm{vol}(\mathcal{D}) = \max_{k,j} |D_{kj}|$. Let $\mathcal{B}_\mathcal{D}$ be the transition matrix of the (heat-bath) block dynamics with blocks $\mathcal{D}$ and let $\boldsymbol{SW}$ denote the transition matrix for the SW dynamics. Then,* $\mathsf{gap}(\boldsymbol{SW}) \ge \exp(-O(\mathrm{vol}(\mathcal{D}))) \cdot \mathsf{gap}(\mathcal{B}_\mathcal{D})$.

The blocks of the tiled block dynamics satisfy all the conditions in this theorem, and, in addition, $\mathrm{vol}(\mathcal{D}) = O(1)$. Hence, combining all the results stated in this section, we see that Theorem 1 from introduction follows.

**Proof of Theorem 1.** Follows from Theorems 9–11. ◀

## 3.1 Equivalence between VM and PVM: Proof of Theorem 9

In this section we establish the equivalence between VM and PVM. We start with the definition of *General Variance Mixing (GVM)*. Let $\Phi$ and $\Psi$ be two finite sets and let $\rho(\cdot,\cdot)$ be an arbitrary joint distribution supported on $\Phi \times \Psi$. Denote by $\nu$ and $\pi$ the marginal distributions of $\rho$ over $\Phi$ and $\Psi$, respectively. That is, for $x \in \Phi$ we have $\nu(x) = \sum_{y\in\Psi} \rho(x,y)$, and for $y \in \Psi$ we have $\pi(y) = \sum_{x\in\Phi} \rho(x,y)$. We consider two natural matrices associated to $\rho$. For $x \in \Phi$ and $y \in \Psi$, define

$$P^\uparrow(x,y) = \rho(y \mid x) = \frac{\rho(x,y)}{\nu(x)}, \text{ and } \qquad P^\downarrow(y,x) = \rho(x \mid y) = \frac{\rho(x,y)}{\pi(y)}; \tag{8}$$

$P^\uparrow$ is a $|\Phi| \times |\Psi|$ matrix while $P^\downarrow$ is a $|\Psi| \times |\Phi|$ matrix. In addition, observe that $P^\uparrow P^\downarrow$ and $P^\downarrow P^\uparrow$ are transition matrices of Markov chains reversible w.r.t. $\nu$ and $\pi$, respectively.

▶ **Definition 12** (GVM for $\rho$)**.** *We say that the joint distribution $\rho$ satisfies* $\mathrm{GVM}(\varepsilon)$ *if for every function $f : \Phi \to \mathbb{R}$ we have $\mathrm{Var}_\pi(P^\downarrow f) \le \varepsilon \cdot \mathrm{Var}_\nu(f)$.*

One key observation in our proof is that the GVM condition can be expressed in term of the spectral gaps of the matrices $P^\uparrow P^\downarrow$ and $P^\downarrow P^\uparrow$.

▶ **Lemma 13.** *The joint distribution $\rho$ satisfies* $\mathrm{GVM}(\varepsilon)$ *if and only if* $\mathsf{gap}(P^\uparrow P^\downarrow) = \mathsf{gap}(P^\downarrow P^\uparrow) \ge 1 - \varepsilon$.

Before providing the proof of Lemma 13, we recall the definition of the *adjoint* operator. Let $S_1$ and $S_2$ be two Hilbert spaces with inner products $\langle \cdot,\cdot \rangle_{S_1}$ and $\langle \cdot,\cdot \rangle_{S_2}$ respectively, and let $K : S_2 \to S_1$ be a bounded linear operator. The adjoint of $K$ is the unique operator $K^* : S_1 \to S_2$ satisfying $\langle f, Kg \rangle_{S_1} = \langle K^* f, g \rangle_{S_2}$ for all $f \in S_1$ and $g \in S_2$. When $S_1 = S_2$, $K$ is called *self-adjoint* if $K = K^*$. We can now provide the proof of Lemma 13.

**Proof of Lemma 13.** It is straightforward to check that $P^\uparrow \mathbf{1} = \mathbf{1}$, $P^\downarrow \mathbf{1} = \mathbf{1}$, $\nu P^\uparrow = \pi$, $\pi P^\downarrow = \nu$, and that the operator $P^\uparrow : L_2(\pi) \to L_2(\nu)$ is the adjoint of the operator $P^\downarrow : L_2(\nu) \to L_2(\pi)$. Hence, both $P^\uparrow P^\downarrow$ and $P^\downarrow P^\uparrow$ are positive semidefinite and have the same multiset of non-zero eigenvalues. Now, for $f : \Phi \to \mathbb{R}$, we have

$$\mathrm{Var}_\pi(P^\downarrow f) = \left\langle P^\downarrow f, (I - \mathbf{1}\pi) P^\downarrow f \right\rangle_\pi = \left\langle f, P^\uparrow (I - \mathbf{1}\pi) P^\downarrow f \right\rangle_\nu = \left\langle f, P^\uparrow P^\downarrow f \right\rangle_\nu - \left\langle f, \mathbf{1}\nu f \right\rangle_\nu .$$

Therefore, $\mathrm{Var}_\pi(P^{\downarrow} f) \le \varepsilon \cdot \mathrm{Var}_\nu(f)$ holds if and only if

$$\begin{aligned} &\left\langle f, P^{\uparrow}P^{\downarrow} f\right\rangle_\nu - \langle f, \mathbf{1}\nu f\rangle_\nu \le \varepsilon \cdot (\langle f, f\rangle_\nu - \langle f, \mathbf{1}\nu f\rangle_\nu) \\ \Leftrightarrow \quad &\left\langle f, (I - P^{\uparrow}P^{\downarrow}) f\right\rangle_\nu \ge (1-\varepsilon)\cdot \langle f, (I-\mathbf{1}\nu) f\rangle_\nu \\ \Leftrightarrow \quad &\mathcal{E}_{P^{\uparrow}P^{\downarrow}}(f,f) \ge (1-\varepsilon)\cdot \mathrm{Var}_\nu(f). \end{aligned}$$

The lemma then follows from (6). ◀

We provide next the proof of Theorem 9, which follows from Lemma 13 and interpretations of VM and PVM by GVM. Given $F = A \cup B \subseteq T$ and $\eta \in \Omega$, let $P^{\uparrow} = (P_F^\eta)_{A\uparrow B}$ denote the $q^{|A\setminus B|} \times q^{|B\setminus A|}$ stochastic matrix indexed by the configurations on the sets $A \setminus B$ and $B \setminus A$, such that for $\xi \in [q]^{A\setminus B}$ and $\xi' \in [q]^{B\setminus A}$ we have $P^{\uparrow}(\xi,\xi') = \mu_F^\eta(\sigma_{B\setminus A} = \xi' \mid \sigma_{A\setminus B} = \xi)$. In words, $P^{\uparrow}$ corresponds to the transition matrix that given the configuration $\xi$ in $A \setminus B$ updates the configuration in $B \setminus A$ from the conditional distribution $\mu_F^\eta(\cdot \mid \xi)$. We define in a similar manner the $q^{|B\setminus A|} \times q^{|A\setminus B|}$ stochastic matrix $P^{\downarrow} = (P_F^\eta)_{B\downarrow A}$ where for $\xi' \in [q]^{B\setminus A}$ and $\xi \in [q]^{A\setminus B}$ we have $P^{\downarrow}(\xi',\xi) = \mu_F^\eta(\sigma_{A\setminus B} = \xi \mid \sigma_{B\setminus A} = \xi')$.

If we set $\rho$ to be the marginal of $\mu_F^\eta$ on $(A \setminus B) \cup (B \setminus A)$, then $\Phi = [q]^{A\setminus B}$, $\Psi = [q]^{B\setminus A}$, and $\nu$ and $\pi$ are the marginals of $\mu_F^\eta$ on $A \setminus B$ and $B \setminus A$, respectively. Therefore, according to Definition 12, GVM($\varepsilon$) holds for the marginal of $\mu_F^\eta$ on $(A \setminus B) \cup (B \setminus A)$ if $\mathrm{Var}_\pi(P^{\downarrow} f) \le \varepsilon \cdot \mathrm{Var}_\nu(f)$ for every function $f : \Phi \to \mathbb{R}$.

Now, note that a function $g : \Omega_F^\eta \to \mathbb{R}$ independent of $B$ only depends on the configuration on $A \setminus B$. Thus, for fixed $\eta$, $g$ induces a function $f : \Phi \to \mathbb{R}$; in particular, $\mathrm{Var}_F^\eta(g) = \mathrm{Var}_\nu(f)$. Moreover, letting $g_{B\setminus A}(\xi) := \mathbb{E}_F^\eta[g \mid \sigma_{B\setminus A} = \xi]$, we have $g_{B\setminus A}(\xi) = P^{\downarrow} f(\xi)$ for every $\xi \in \Psi = [q]^{B\setminus A}$, and so $\mathrm{Var}_F^\eta(g_{B\setminus A}) = \mathrm{Var}_\pi(P^{\downarrow} f)$. Consequently, we arrive at the following equivalences between VM, PVM and GVM.

▶ **Proposition 14.**
1. *The Gibbs distribution $\mu$ satisfies* VM($\ell,\epsilon$) *if and only if for every $v \in T$ and $\eta \in \Omega$,* GVM($\varepsilon$) *holds for the marginal of $\mu_{T_v}^\eta$ on $(T_v \setminus B(v,\ell)) \cup \{v\}$.*
2. *The Gibbs distribution $\mu$ satisfies* PVM($\ell,\epsilon$) *if and only if for every $i$ such that $1 \le i \le h+1$ and $\eta \in \Omega$,* GVM($\varepsilon$) *holds for the marginal of $\mu_{F_i}^\eta$ on $(F_i \setminus \cup_{v\in L_i} B(v,\ell)) \cup L_i$.*

To see part 1 simply note that in the notation above, we can set $F = T_v$, $A = T_v \setminus v$ and $B = B(v,\ell)$. For part 2, we set $F = F_i$, $A = F_{i-1}$ and $B = B_i^\ell$.

**Proof of Theorem 9.** From Proposition 14 and Lemma 13, VM($\ell,\epsilon$) holds if and only if $\mathsf{gap}(Q_v) \ge 1-\varepsilon$ for every $v \in T$ and $\eta \in \Omega$, where $Q_v = (P_{T_v}^\eta)_{B(v,\ell)\downarrow(T_v\setminus v)}(P_{T_v}^\eta)_{(T_v\setminus v)\uparrow B(v,\ell)}$. Similarly, $\mu$ satisfies PVM($\ell,\epsilon$) if and only if $\mathsf{gap}(Q_{L_i}) \ge 1-\varepsilon$ for every $i$ such that $1 \le i \le h+1$ and $\eta \in \Omega$, where $Q_{L_i} = (P_{F_i}^\eta)_{B_i^\ell\downarrow F_{i-1}}(P_{F_i}^\eta)_{F_{i-1}\uparrow B_i^\ell}$.

Since $F_i = \bigcup_{v\in L_i} T_v$ and the $T_v$'s are at distance at least two from each other, $\mu_{F_i}^\eta(\sigma_{L_i} = \cdot)$ is a product distribution; in particular $\mu_{F_i}^\eta(\sigma_{L_i} = \cdot) = \prod_{v\in L_i} \mu_{T_v}^\eta(\sigma_v = \cdot)$ and the chain with transition matrix $Q_{L_i}$ is a product Markov chain where each component corresponds to $Q_v$ for some $v \in L_i$. A standard fact about product Markov chains, see, e.g., [9, Lemma 4.7], then implies that $\mathsf{gap}(Q_{L_i}) = \min_{v\in L_i} \mathsf{gap}(Q_v)$ and the result follows. ◀

## 4 Entropy Mixing: Proof of Theorem 2

Let $E \subseteq T$ denote the set of all *even* vertices of the tree $T$, where a vertex is called even if its distance to the leaves is even; let $O = T \setminus E$ be the set of all odd vertices. We show that EM (i.e., entropy mixing) as defined in [40] implies a factorization of entropy into even and odd subsets of vertices. This even-odd factorization was recently shown to imply $O(\log n)$ mixing of the SW dynamics on bipartite graphs [8].

We start with the definition of EM, which is the analog of the VM condition for entropy. Let $\tau$ be a fixed boundary condition and again set $\mu := \mu_T^\tau$ and $\Omega := \Omega^\tau$ for ease of notation. Recall that for $v \in T$, we use $T_v$ for the subtree of $T$ rooted at $v$. Recall that for $\eta \in \Omega$ and $g : \Omega_{T_v}^\eta \to \mathbb{R}$, we defined the function $g_v(a) = \mathbb{E}_{T_v}^\eta[g \mid \sigma_v = a]$ for $a \in [q]$; see (7).

▶ **Definition 15** (Entropy Mixing (EM)). *The Gibbs distribution $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, \varepsilon)$ if for every $v \in T$, every $\eta \in \Omega$, and every function $g : \Omega_{T_v}^\eta \to \mathbb{R}$ that is independent of the configuration on $B(v, \ell)$, we have $\mathrm{Ent}_{T_v}^\eta(g_v) \le \varepsilon \cdot \mathrm{Ent}_{T_v}^\eta(g)$. The EM condition holds if there exist constants $\ell$ and $\varepsilon = \varepsilon(\ell)$ such that $\mathrm{EM}(\ell, \varepsilon)$ holds.*

Extending our notation from the previous section for the variance functional, for $A \subseteq T$ and a function $f : \Omega \to \mathbb{R}_{\ge 0}$, we use $\mathrm{Ent}_A(f)$ for the conditional entropy of $f$ w.r.t. $\mu$ given a spin configuration in $T \setminus A$; i.e., for $\xi \in \Omega$ we have

$$(\mathrm{Ent}_A(f))(\xi) = \mathrm{Ent}_A^\xi(f) = \mathrm{Ent}_\mu[f \mid \sigma_{T \setminus A} = \xi_{T \setminus A}].$$

In particular, we shall write $\mathrm{Ent}(f) = \mathrm{Ent}_T(f) = \mathrm{Ent}_\mu(f)$. Notice that $\mathrm{Ent}_A(f)$ can be viewed as a function from $[q]^{T \setminus A}$ to $\mathbb{R}_{\ge 0}$ and $\mathbb{E}[\mathrm{Ent}_A(f)]$ denotes its mean, averaging over the configuration on $T \setminus A$. We state next our even-odd factorization of entropy.

▶ **Theorem 16.** *If there exist $\ell \in \mathbb{N}^+$ and $\varepsilon \in (0,1)$ such that $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, \varepsilon)$, then there exists a constant $C_{\text{EO}} = C_{\text{EO}}(\ell, \varepsilon)$ independent of $n$ such that for every function $f : \Omega \to \mathbb{R}_{\ge 0}$ we have $\mathrm{Ent}(f) \le C_{\text{EO}} \left( \mathbb{E}[\mathrm{Ent}_E(f)] + \mathbb{E}[\mathrm{Ent}_O(f)] \right)$.*

Theorem 2 follows immediately.

**Proof of Theorem 2.** By Theorem 16, EM implies the even-odd factorization of entropy, and the results in [8] imply that the mixing time of the SW dynamics is $O(\log n)$. ◀

Our main technical contribution in the proof Theorem 2 is thus Theorem 16; namely, that EM implies the even-odd factorization of entropy. To prove Theorem 16, we will first establish entropy factorization for the tiled blocks defined in (3) and (2); see also Figures 1(b) and 1(c). From the tiled block factorization of entropy we then deduce the desired even-odd factorization. This approach is captured by the following two lemmas.

▶ **Lemma 17.** *If there exist $\ell \in \mathbb{N}^+$ and $\varepsilon \in (0,1)$ such that $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, \varepsilon)$, then there exists a constant $C_{\text{TB}} = C_{\text{TB}}(\ell, \varepsilon)$ independent of $n$ such that, for every function $f : \Omega \to \mathbb{R}_{\ge 0}$, $\mathrm{Ent}(f) \le C_{\text{TB}} \cdot \sum_{j=1}^{\ell+1} \mathbb{E}[\mathrm{Ent}_{T_j^\ell}(f)]$.*

▶ **Lemma 18.** *If for every function $f : \Omega \to \mathbb{R}_{\ge 0}$ we have $\mathrm{Ent}(f) \le C_{\text{TB}} \cdot \sum_{j=1}^{\ell+1} \mathbb{E}[\mathrm{Ent}_{T_j^\ell}(f)]$, then there exists $C_{\text{EO}} = C_{\text{EO}}(C_{\text{TB}}, \ell)$ such that for every function $f : \Omega \to \mathbb{R}_{\ge 0}$ we have*

$$\mathrm{Ent}(f) \le C_{\text{EO}} \left( \mathbb{E}[\mathrm{Ent}_E(f)] + \mathbb{E}[\mathrm{Ent}_O(f)] \right).$$

**Proof of Theorem 16.** Follows directly from Lemmas 17 and 18. ◀

We proved a version of Lemma 17 for the variance functional as part of the proof of Theorem 10, and the same argument can then be easily adapted to entropy. We provide next the proof of Lemma 18, which contains the main novelty in our proof of Theorem 16.

**Proof of Lemma 18.** First, we claim that there exists a constant $C' = C'(\ell)$ such that for every function $f : \Omega^{\eta}_{B(v,\ell)} \to \mathbb{R}_{\geq 0}$ one has the following inequality:

$$\mathrm{Ent}^{\eta}_{B(v,\ell)}(f) \leq C' \left( \mathbb{E}^{\eta}_{B(v,\ell)}[\mathrm{Ent}_{B(v,\ell)\cap E}(f)] + \mathbb{E}^{\eta}_{B(v,\ell)}[\mathrm{Ent}_{B(v,\ell)\cap O}(f)] \right). \tag{9}$$

To deduce (9), consider the even-odd block dynamics $M$ in $B(v,\ell)$ with boundary condition $\eta$ and blocks $\mathcal{U} = \{E \cap B(v,\ell), O \cap B(v,\ell)\}$. A simple coupling argument implies that the spectral gap of $M$ is $\Omega(1)$. Then, Corollary A.4 from [19] implies that the log-Sobolev constant $\alpha(M)$ of $M$ is $\Omega(1)$, which establishes (9) with constant $C' = O(1/\alpha(M))$. We note that all bounds and comparisons in this argument are fairly crude, and, in fact, the constant $C'$ depends exponentially on $|B(v,\ell)|$, but it is still independent of $n$.

Next, notice that, for any $\eta \in \Omega$, $\mu^{\eta}_{T^{\ell}_j}$ is the product of a collection of distributions on (disjoint) subsets $B(v,\ell)$. Lemma 3.2 from [14] allows us to lift the "local" even-odd factorization in each $B(v,\ell)$ from (9) to a "global" even-odd factorization in $T^{\ell}_j$. Specifically, for every function $f : \Omega^{\eta}_{T^{\ell}_j} \to \mathbb{R}_{\geq 0}$ we obtain

$$\mathrm{Ent}^{\eta}_{T^{\ell}_j}(f) \leq C' \left( \mathbb{E}^{\eta}_{T^{\ell}_j}[\mathrm{Ent}_{T^{\ell}_j\cap E}(f)] + \mathbb{E}^{\eta}_{T^{\ell}_j}[\mathrm{Ent}_{T^{\ell}_j\cap O}(f)] \right).$$

Taking expectation over $\eta$, we get

$$\mathbb{E}[\mathrm{Ent}_{T^{\ell}_j}(f)] \leq C' \left( \mathbb{E}[\mathrm{Ent}_{T^{\ell}_j\cap E}(f)] + \mathbb{E}[\mathrm{Ent}_{T^{\ell}_j\cap O}(f)] \right) \leq C' \left( \mathbb{E}[\mathrm{Ent}_E(f)] + \mathbb{E}[\mathrm{Ent}_O(f)] \right);$$

the last inequality follows from the fact that $\mathrm{Ent}^{\eta}_E(f) = \mathbb{E}^{\eta}_E[\mathrm{Ent}_{T^{\ell}_j\cap E}(f)] + \mathrm{Ent}^{\eta}_E[\mathbb{E}_{T^{\ell}_j\cap E}(f)]$. Summing up over $j$,

$$\sum_{j=1}^{\ell+1} \mathbb{E}[\mathrm{Ent}_{T^{\ell}_j}(f)] \leq C'(\ell+1) \left( \mathbb{E}[\mathrm{Ent}_E(f)] + \mathbb{E}[\mathrm{Ent}_O(f)] \right),$$

and the result follows by taking $C_{\text{EO}} = C'(\ell+1)$. ◀

## References


1 Blanca A., Chen Z., Štefankovič D., and Vigoda E. The Swendsen-Wang dynamics on trees. *arXiv preprint arXiv:2007.08068*, 2020.

2 V. L. Alev and L. C. Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2020.

3 N. Anari, K. Liu, and S. Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, 2020.

4 B. Awerbuch and Y. Shiloach. New connectivity and MSF algorithms for shuffle-exchange network and PRAM. *IEEE Computer Architecture Letters*, 36(10):1258–1263, 1987.

5 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.

6 N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

7 H. A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 150(871):552–575, 1935.

8 A. Blanca, P. Caputo, D. Parisi, A. Sinclair, and E. Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$. In *Proceedings of the 53st Annual ACM Symposium on Theory of Computing (STOC)*, page 1551–1564, 2021.

9 A. Blanca, P. Caputo, A. Sinclair, and E. Vigoda. Spatial Mixing and Non-local Markov chains. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1965–1980, 2018.

10 A. Blanca, A. Galanis, L. A. Goldberg, D. Štefankovič, E. Vigoda, and K. Yang. Sampling in uniqueness from the Potts and random-cluster models on random regular graphs. *SIAM Journal on Discrete Mathematics*, 34(1):742–793, 2020.

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

12 A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. *Proceedings of the 19th International Workshop on Randomization and Computation*, pages 528–543, 2015.

13 M. Bordewich, C. Greenhill, and V. Patel. Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Structures & Algorithms*, 48(1):21–52, 2016.

14 P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing, 2020. URL: `https://arxiv.org/abs/2004.10574`.

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 C. Cooper and A. M. Frieze. Mixing properties of the Swendsen-Wang process on classes of graphs. *Random Structures and Algorithms*, 15(3-4):242–261, 1999.

17 M. Costeniuc, R. S. Ellis, and H. Touchette. Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model. *Journal of Mathematical Physics*, 46(6):063301, 2005.

18 P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012.

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

20 H. Duminil-Copin, M. Gagnebin, M. Harel, I. Manolescu, and V. Tassion. Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. *Annales de l'ENS*, 2016.

21 H. Duminil-Copin, V. Sidoravicius, and V. Tassion. Continuity of the Phase Transition for Planar Random-Cluster and Potts Models with $1 \leq q \leq 4$. *Communications in Mathematical Physics*, 349(1):47–107, 2017.

22 M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structure & Algorithms*, 24(4):461–479, 2004.

23 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.

24 Grimmett G. The random-cluster model. In *Probability on discrete structures*, pages 73–123. Springer, 2004.

25 A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. In *Proceedings of the 19th International Workshop on Randomization and Computation*, pages 815–828, 2015.

26 A. Galanis, D. Štefankovič, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016.

27 H. O. Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter Studies in Mathematics. Walter de Gruyter Inc, 1988.

28 A. Gerschenfeld and A. Montanari. Reconstruction for models on random graphs. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 194–204, 2007.

29 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.

30 D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

**31** O. Häggström. The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104:231–253, 1996.

**32** T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 511–520, 2005.

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

**34** M. Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics, Birkhäuser Verlag, 2003.

**35** M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51(4):671–697, 2004.

**36** J. Jonasson. The random cluster model on a general graph and a phase transition characterization of nonamenability. *Stochastic Processes and their Applications*, 79(2):335–354, 1999.

**37** R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random structures and algorithms*, 11(1):1–50, 1997.

**38** D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

**39** F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. *Communications in Mathematical Physics*, 161(3):447–486, 1994.

**40** F. Martinelli, A. Sinclair, and D. Weitz. The Ising model on trees: Boundary conditions and mixing time. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 628–639, 2003.

**41** F. Martinelli, A. Sinclair, and D. Weitz. Fast mixing for independent sets. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 449–458. colorings and other model on trees. In, 2004.

**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** E. Mossel and Y. Peres. Information flow on trees. *Annals of Applied Probability*, 13(3):817–844, 2003.

**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** R. Restrepo, D. Štefankovič, C. Vera, E. Vigoda, and L. Yang. Phase transition for Glauber dynamics for independent sets on regular trees. *SIAM Journal on Discrete Mathematics*, 28(2):835–861, 2014.

**46** L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics*, pages 301–413. 1997.

**47** A. Sly. Reconstruction for the Potts model. *The Annals of Probability*, 39(4):1365–1406, 2011.

**48** A. Sly and Y. Zhang. The Glauber dynamics of colorings on trees is rapidly mixing throughout the nonreconstruction regime. *The Annals of Applied Probability*, 27(5):2646–2674, 2017.

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

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