Rapid mixing of global Markov chains via spectral independence: the unbounded degree case

Antonio Blanca [*] Xusheng Zhang [†]

August 30, 2023

Abstract

We consider spin systems on general n-vertex graphs of unbounded degree and explore the effects of spectral independence on the rate of convergence to equilibrium of global Markov chains. Spectral independence is a novel way of quantifying the decay of correlations in spin system models, which has significantly advanced the study of Markov chains for spin systems. We prove that whenever spectral independence holds, the popular Swendsen–Wang dynamics for the q-state ferromagnetic Potts model on graphs of maximum degree Δ, where Δ is allowed to grow with n, converges in $O((\Delta \log n)^c)$ steps where $c > 0$ is a constant independent of Δ and n. We also show a similar mixing time bound for the block dynamics of general spin systems, again assuming that spectral independence holds. Finally, for *monotone* spin systems such as the Ising model and the hardcore model on bipartite graphs, we show that spectral independence implies that the mixing time of the systematic scan dynamics is $O(\Delta^c \log n)$ for a constant $c > 0$ independent of Δ and n. Systematic scan dynamics are widely popular but are notoriously difficult to analyze. Our result implies optimal $O(\log n)$ mixing time bounds for any systematic scan dynamics of the ferromagnetic Ising model on general graphs up to the tree uniqueness threshold. Our main technical contribution is an improved factorization of the entropy functional: this is the common starting point for all our proofs. Specifically, we establish the so-called k-partite factorization of entropy with a constant that depends polynomially on the maximum degree of the graph.

[*]Pennsylvania State University. Email: ablanca@cse.psu.edu. Research supported in part by NSF grant CCF-2143762.
[†]Pennsylvania State University. Email: xushengz@psu.edu. Research supported in part by NSF grant CCF-2143762.

1 Introduction

Spectral independence is a powerful new approach for quantifying the decay of correlations in spin system models. Initially introduced in [ALOG20], this condition has revolutionized the study of Markov chains for spin systems. In a series of important and recent contributions, spectral independence has been shown to be instrumental in determining the convergence rate of the Glauber dynamics, the simple single-site update Markov chain that updates the spin at a randomly chosen vertex in each step.

The first efforts in this series (see [ALOG20,CLV20,CLV21]) showed that spectral independence implies optimal $O(n \log n)$ mixing of the Glauber dynamics on n-vertex graphs of bounded degree for general spin systems. The unbounded degree case was studied in [CFYZ22b,CFYZ22a,AJK+22,JPV22], while [BCC+22] explored the effects of this condition on the speed of convergence of global Markov chains (i.e., Markov chains that update the spins of a large number of vertices in each step) in the bounded degree setting. Research exploring the applications of spectral independence is ongoing. We contribute to this line of work by investigating how spectral independence affects the speed of convergence of *global Markov chains* for general spin systems on graphs of unbounded degree.

A *spin system* is defined on a graph $G = (V, E)$. There is a set $\mathcal{S} = \{1, \dots, q\}$ of spins or colors, and configurations are assignments of spin values from $\mathcal{S}$ to each vertex of G. The probability of a configuration $\sigma \in \mathcal{S}^V$ is given by the Gibbs distribution:

$$\mu(\sigma) = \frac{e^{-H(\sigma)}}{Z}, \tag{1}$$

where the normalizing factor Z is known as the partition function, and the Hamiltonian $H : \mathcal{S}^V \to \mathbb{R}$ contains terms that depend on the spin values at each vertex (a "vertex potential" or "external field") and at each pair of adjacent vertices (an "edge potential"); see Definition 2.1. A widely studied spin system, and one that we will pay close attention to in this paper, is the ferromagnetic Potts model, where for a real parameter $\beta > 0$, associated with inverse temperature in physical applications, the Hamiltonian is given by:

$$H(\sigma) = -\beta \sum_{\{u,v\} \in E} \mathbb{1}(\sigma_u = \sigma_v).$$

The classical ferromagnetic Ising model corresponds to the $q = 2$ case. (In this variant of the Potts model, the Hamiltonian only includes edge potentials, and there is no external field.) We shall use μ_{Ising} and μ_{Potts} for the Gibbs distributions corresponding to the Ising and Potts models. Other well-known, well-studied spin systems include uniform proper colorings and the hardcore model.

Spin systems provide a robust framework for studying interacting systems of simple elements and have a wide range of applications in computer science, statistical physics, and other fields. In such applications, generating samples from the Gibbs distribution (1) is a fundamental computational task and one in which Markov chain-based algorithms have been quite successful. A long line of work dating back to the 1980s relates the speed of convergence of Markov chains to various forms of decay of correlations in the model. Spectral independence, defined next, captures the decay of correlations in a novel way.

Roughly speaking, spectral independence holds when the spectral norm of a "pairwise" influence matrix is bounded. To formally define it, let us begin by introducing some notations. Let $\Omega \subseteq \mathcal{S}^V$ be the support of μ: the set of configurations σ such that $\mu(\sigma) > 0$. A *pinning* τ on a subset of vertices $\Lambda \subseteq V$ is a fixed partial configuration on Λ; i.e., a spin assignment from $\mathcal{S}^\Lambda$ to the vertices of Λ. For a pinning τ on $\Lambda \subseteq V$ and $U \subseteq V \setminus \Lambda$, we let $\Omega_U^\tau = \{\sigma_U \in \mathcal{S}^U : \mu(\sigma_U \mid \sigma_\Lambda = \tau) > 0\}$ be the set of partial configurations on U that are consistent with the pinning τ. We write $\Omega_u^\tau = \Omega_{\{u\}}^\tau$ if u is a single vertex. Let

$$\mathcal{P}^\tau := \{(u, s) : u \notin \Lambda, s \in \Omega_u^\tau\}$$

denote the set of consistent vertex-spin pairs in $\Omega^{\tau}_{V\setminus\Lambda}$ under μ. For each $\Lambda \subseteq V$ and pinning τ on Λ, we define the *signed pairwise influence matrix* $\Psi^{\tau}_{\mu} \in \mathbb{R}^{\mathcal{P}^{\tau}\times\mathcal{P}^{\tau}}$ to be the matrix with entries:

$$\Psi^{\tau}_{\mu}((u,a),(v,b)) = \mu(\sigma_v = b \mid \sigma_u = a, \sigma_\Lambda = \tau) - \mu(\sigma_v = b \mid \sigma_\Lambda = \tau)$$

for $u \neq v$, and $\Psi^{\tau}_{\mu}((u,a),(u,b)) = 0$ otherwise.

Definition 1.1 (Spectral Independence)**.** A distribution μ satisfies *η-spectral independence* if for every subset of vertices $\Lambda \subseteq V$ and every pinning $\tau \in \Omega_\Lambda$, the largest eigenvalue of the signed pairwise influence matrix Ψ^{τ}_{μ}, denoted $\lambda_1(\Psi^{\tau}_{\mu})$, satisfies $\lambda_1(\Psi^{\tau}_{\mu}) \leq \eta$.

There are several definitions of spectral independence in the literature; we use here the one from [CGSV21].

We show that spectral independence implies new upper bounds on the mixing time of several well-studied global Markov chains in the case where the maximum degree Δ of the underlying graph $G = (V, E)$ is unbounded; i.e., $\Delta \to \infty$ with n. The mixing time is defined as the number of steps required for a Markov chain to reach a distribution close in total variation distance to its stationary distribution, assuming a worst possible starting state; a formal definition is given in Section 2.1. The global Markov chains we consider include the Swendsen–Wang dynamics for the ferromagnetic q-state Potts, the systematic scan dynamics for monotone spin systems, and the block dynamics for general spin systems. These three dynamics are among the most popular and well-studied global Markov chains and present certain advantages (e.g., faster convergence and amenability to parallelization) to the Glauber dynamics.

1.1 The Swendsen–Wang dynamics

A canonical example of a global Markov chain is the Swendsen–Wang (SW) dynamics for the ferromagnetic q-state Potts model. The SW dynamics transitions from a configuration σ_t to σ_{t+1} by:

1. For each edge $e = \{u, v\} \in E$, if $\sigma_t(u) = \sigma_t(v)$, independently include e in the set A_t with probability $p = 1 - e^{-\beta}$;
2. Then, independently for each connected component C in (V, A_t), draw a spin $s \in \{1, \dots, q\}$ uniformly at random and set $\sigma_{t+1}(v) = s$ for all $v \in C$.

The SW dynamics is ergodic and reversible with respect to μ_{Potts} and thus converges to it. This Markov chain originated in the late 1980s [SW87] as an alternative to the Glauber dynamics, which mixes exponentially slowly at low temperatures (large β). The SW dynamics bypasses the key barriers that cause the slowdown of the Glauber dynamics at low temperatures. For the Ising model ($q = 2$), for instance, it was recently shown to converge in poly(n) steps on any n-vertex graph for any value of $\beta > 0$ [GJ17]. (The conjectured mixing time is $\Theta(n^{1/4})$, but we seem to be far from proving such a conjecture.) For $q \geq 3$, on the other hand, the SW dynamics can converge exponentially slowly at certain "intermediate" temperatures regimes corresponding to first-order phase transitions; see [GJ97, BCT12, GL18, GLP19, COGG+23].

Recently, η-spectral independence (with $\eta = O(1)$) was shown to imply that the mixing time of the SW dynamics is $O(\log n)$ on graphs of maximum degree $\Delta = O(1)$, i.e., bounded degree graphs [BCC+22]. This mixing time bound is optimal since the SW dynamics requires $\Omega(\log n)$ steps to mix in some cases where η and Δ are both $O(1)$ [BCP+22, BCSV23]. However, it does not extend to the unbounded degree setting since the constant factor hidden by the big-O notation depends exponentially on the maximum degree Δ; this is the case even when $\eta = O(1)$ and $\beta\Delta = O(1)$. Our first result provides a mixing time bound that depends only polynomially on Δ.

Theorem 1.2. *Let $q \geq 2$, $\beta > 0$, $\eta > 0$ and $\Delta \geq 3$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ. Let μ_{Potts} be the Gibbs distribution of the q-state ferromagnetic Potts model on G with parameter β.*

If μ_{Potts} is η-spectrally independent with $\eta = O(1)$ and $\beta\Delta = O(1)$, then there exists a constant $c > 0$ such that the mixing time of the SW dynamics satisfies $T_{mix}(P_{SW}) = O\big((\Delta \log n)^c\big)$.

The constant c has a near linear dependency on η and $\beta\Delta$; a more precise statement of Theorem 1.2 with a precise expression for c is given in Theorem 3.1.

Despite the expectation that the SW dynamics mixes in $O(\log n)$ steps in weakly correlated systems (i.e., when $\beta\Delta$ is small), proving sub-linear upper bounds on its mixing time has been difficult. Recently, various forms of decay of correlation (e.g., strong spatial mixing, entropy mixing, and spectral independence) have been used to obtain $O(\log n)$ bounds for the mixing time of the SW dynamics on cubes of the integer lattice graph $\mathbb{Z}^d$, regular trees, and general graphs of bounded degree (see [BCP+22, BCSV23, BCC+22]). However, for graphs of large degree, i.e., with $\Delta \to \infty$ with n, the only sub-linear mixing time bounds known either hold for the very distinctive mean-field model, where G is the complete graph [GŠV15, BS15], or hold for very small values of β; i.e., $\beta \lesssim 1/(3\Delta)$ [Hub03]. Our results provide new sub-linear mixing time bounds for graph families of sub-linear maximum degree, provided $\eta = O(1)$ and $\beta\Delta = O(1)$. These last two conditions go hand-in-hand: in all known cases where $\eta = O(1)$, we also have $\beta\Delta = O(1)$.

On graphs of degree at most Δ, η-spectral independence is supposed to hold with $\eta = O(1)$ whenever $\beta < \beta_u(q, \Delta)$, where $\beta_u(q, \Delta)$ is the threshold for the uniqueness/non-uniqueness phase transition on Δ-regular trees. This has been confirmed for the Ising model ($q = 2$) but not for the Potts model. Specifically, for the ferromagnetic Ising model, we have $\beta_u(2, \Delta) = \ln \frac{\Delta}{\Delta-2}$, and when $\beta \le (1-\delta)\beta_u(2, \Delta)$ for some $\delta \in (0, 1)$, μ_{Ising} is η-spectrally independent with $\eta = O(1/\delta)$; see [CLV20, CLV21]. In contrast, for the ferromagnetic Potts model with $q \ge 3$, there is no closed-form expression for $\beta_u(q, \Delta)$ (it is defined as the threshold value where an equation starts to have a double root), and for graphs of unbounded degree η-spectral independence is only known to hold when $\beta \le \frac{2(1-\delta)}{\Delta}$. As a result, we obtain the following corollary of Theorem 1.2.

Corollary 1.3. *Let $\delta \in (0, 1)$, $\Delta \ge 3$. Suppose that either $q = 2$ and $\beta < (1-\delta)\beta_u(2, \Delta)$, or $q \ge 3$ and $\beta \le \frac{2(1-\delta)}{\Delta}$. Then, there exists a constant $c = c(\delta) > 0$ such that the mixing time of the SW dynamics for the q-state ferromagnetic Potts model on any n-vertex graph of maximum degree Δ satisfies $T_{mix}(P_{SW}) = O\big((\Delta \log n)^c\big)$.*

We mention that other conditions known to imply spectral independence (e.g., those in [BGP16]) are not well-suited for the unbounded degree setting since under those conditions, the best known bound for η depends polynomially on Δ. For another application of Theorem 1.2, see Section 3.5.1 where we provide a bound on the mixing of the SW dynamics on random graphs.

We comment briefly on our proof approach for Theorem 1.2. A mixing time bound for the SW dynamics can be deduced from the so-called *edge-spin* factorization of the entropy functional introduced in [BCP+22]. It was noted there that this factorization, in turn, follows from a different factorization of entropy known as *k-partite factorization*, or KPF. Spectral independence is known to imply KPF but with a loss of a multiplicative constant that depends exponentially on the maximum degree of the graph. Our proof of Theorem 1.2 follows this existing framework, but pays closer attention to establishing KPF with an optimized constant with a better dependence on the model parameters. This is done through a multi-scale analysis of the entropy functional; in each scale, we apply spectral independence to achieve a tighter KPF condition. Our new results for KPF not only hold for the Potts model, but also for a general class of spin systems, and we use it to establish new mixing time bounds for the systematic scan and block dynamics.

1.2 The systematic scan dynamics

Our next contribution pertains the *systematic scan dynamics*, which is a family of Markov chains closely related to the Glauber dynamics in the sense that updates occur at single vertices sequentially. The key difference is that the vertex updates happen according to a predetermined ordering ϕ of the vertices instead of at random vertices. These dynamics offer practical advantages since there is no need to randomly select vertices at each step, thereby reducing computation time. Throughout the paper, we will consider the *heat-bath* vertex updates in which a new spin is assigned to a vertex by sampling from the conditional distribution at the vertex given the spins of its neighbors; this will be the case for both the Glauber and systematic scan dynamics.

There is a folklore belief that the mixing time of the systematic scan dynamics (properly scaled) is closely related to that of the Glauber dynamics. However, analyzing this type of dynamics has proven very challenging (see, e.g., [DGJ06a, Hay06, DGJ09, DGJ06b, PW13, GKZ18, BCSV19]), and the best general condition under which the systematic scan dynamics is known to be optimally mixing is a Dobrushin-type condition due to Dyer, Goldberg, and Jerrum [DGJ09]. The new developments on Markov chain mixing stemming from spectral independence have not yet provided new results for this dynamics, even for the bounded degree case where much progress has already been made. We show that spectral independence implies optimal mixing of the systematic scan dynamics for *monotone* spin systems with *bounded marginals*; we define both of these notions next.

Definition 1.4 (Monotone spin system)**.** In a monotone system, there is a linear ordering of the spins at each vertex which induces a partial order $\preceq_q$ over the state space. A spin system is *monotone* with respect to the partial order $\preceq_q$ if for every $\Lambda \subseteq V$ and every pair of pinnings $\tau_1 \succeq_q \tau_2$ on $V \setminus \Lambda$, the conditional distribution $\mu(\cdot \mid \sigma_\Lambda = \tau_1)$ stochastically dominates $\mu(\cdot \mid \sigma_\Lambda = \tau_2)$.

Canonical examples of monotone spin systems include the ferromagnetic Ising model and the hardcore model on bipartite graphs. As in earlier work (see [CLV20, CLV21, BCC+22]), our bounds on the mixing time will depend on a lower bound on the marginal probability of any vertex-spin pair. This is formalized as follows.

Definition 1.5 (Bounded marginals)**.** The distribution μ is said to be *b-marginally bounded* if for every $\Lambda \subseteq V$ and pinning $\tau \in \Omega_\Lambda$, and each $(v, s) \in \mathcal{P}^\tau$, we have $\mu(\sigma_v = s \mid \sigma_\Lambda = \tau) \geq b$.

Before stating our result for the systematic scan dynamics of b-marginally bounded monotone spin systems, we note that this Markov chain updates in a single step each vertex once in the order prescribed by ϕ. Under a minimal assumption on the spin system (the same one required to ensure the ergodicity of the Glauber dynamics), the systematic scan dynamics is ergodic. Specifically, when the spin system is totally-connected (see Definition 2.2), the systematic scan dynamics is ergodic. Moreover, the systematic scan dynamics is not necessarily reversible with respect to μ, so, as in earlier works, we work with the symmetrized version of the dynamics in which, in each step, the vertices are updated according to ϕ first, and subsequently in the reverse order of ϕ. The resulting dynamics, which we denote by P_ϕ, is reversible with respect to μ. Our main result for the systematic scan dynamics is the following.

Theorem 1.6. *Let $b > 0$, $\eta > 0$, and $\Delta \geq 3$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ. Let μ be the distribution of a totally-connected monotone spin system on G. If μ is η-spectrally independent and b-marginally bounded, then there exists a universal constant $C > 0$ such that for any ordering ϕ*

$$T_{mix}(P_\phi) = \Delta^{9+4\lceil \frac{2\eta}{b} \rceil} \cdot \left(\frac{C(\eta+1)^5}{b^6} \right)^{2+\lceil \frac{2\eta}{b} \rceil} \cdot O(\log n).$$

The bound in this theorem is tight: for a particular ordering ϕ, we prove an $\Omega(\log n)$ mixing time lower bound that applies to settings where Δ, b and η are all $\Theta(1)$; see Lemma 4.1.

We present next several interesting consequences of Theorem 1.6. First, we obtain the following corollary using the known results about spectral independence for the ferromagnetic Ising model.

Corollary 1.7. *Let $\delta \in (0,1), \Delta \geq 3$ and $0 < \beta < (1-\delta)\beta_u(2,\Delta)$. Suppose $G = (V,E)$ is an n-vertex graph of maximum degree Δ. For any ordering ϕ of the vertices of G, the mixing time of P_ϕ for the Ising model on G with parameter β satisfies $T_{mix}(P_\phi) = O(\log n)$.*

The constant hidden by the big-O notation is an absolute constant that depends only on the constant δ, even when Δ depends on n. This result, compared to the earlier conditions in [DGJ06a, Hay06, DGJ09], extends the parameter regime where the $O(\log n)$ mixing time bound applies; in fact, the parameter regime in Corollary 1.7 is tight, as the systematic scan dynamics undergoes an exponential slowdown when $\beta > \beta_u(2,\Delta)$ [PW13]. We also derive results for the hardcore model on bipartite graphs; see Section 4.3.

Our next application concerns the specific but relevant case where the underlying graph is an n-vertex cube of the integer lattice graph $\mathbb{Z}^d$. In this context, it was proved in [BCSV19] that all systematic scan dynamics converge in $O(\log n(\log\log n)^2)$ steps whenever a well-known condition known as *strong spatial mixing (SSM)* holds. A pertinent open question is whether SSM implies spectral independence. In fact, spectral independence is often proved by adapting earlier arguments for establishing SSM (see, e.g., [ALOG20, CLV20]). Recently, it was proved in [CLMM23] that SSM on trees implies spectral independence on large-girth graphs. We show that for *general* spin systems on $\mathbb{Z}^d$, SSM implies η-spectral independence with $\eta = O(1)$.

Lemma 1.8. *For a spin system on a d-dimensional cube $V \subseteq \mathbb{Z}^d$, SSM implies η-spectral independence, where $\eta = O(1)$.*

The formal definition of SSM is given later in Section 4. Lemma 1.8 does not assume monotonicity for the spin system and could be of independent interest. An interesting consequence of this lemma, when combined with Theorem 1.6 is the following.

Corollary 1.9. *Let $d \geq 2$ and $b > 0$. For a b-marginally bounded monotone totally-connected spin system on a d-dimensional cube $V \subseteq \mathbb{Z}^d$, SSM implies that the mixing time of any systematic scan P_ϕ is $O(\log n)$.*

For the ferromagnetic Ising model on $\mathbb{Z}^2$, SSM is known to hold for all $\beta < \beta_c(2) = \ln(1+\sqrt{2})$ (see [CP21, MOS94, Ale98, BDC12]), so by Corollary 1.9 we deduce that when $\beta < \beta_c(2)$, the mixing time of any systematic scan P_ϕ on an n-vertex square box of $\mathbb{Z}^2$ is $O(\log n)$; note that $\beta_c(2) > \beta_u(2,2d)$, the corresponding tree uniqueness threshold.

We comment briefly on the techniques used to establish our results for the systematic scan dynamics. Our starting point is again the k-partite factorization of entropy (KPF). Our improved bounds for KPF imply that a global Markov chain that updates a random independent set of vertices in each step is rapidly mixing. We then use the censoring technique from [FK13, BCV20] to relate the mixing time of this Markov chain to that of the systematic scan dynamics. To establish Lemma 1.8, we use SSM to construct a contractive coupling for a particular Markov chain. Our Markov chain is similar to the one from [DSVW04], but modified to update rectangles instead of balls, and thus match the variant of SSM that holds up to the critical threshold for the Ising model on $\mathbb{Z}^2$. This contractive coupling is then used to establish spectral independence using the machinery from [BCC+22].

1.3 The block dynamics

Our final result concerns a family of Markov chains known as the *block dynamics*. They are a natural generalization of the Glauber dynamics where a random subset of vertices (instead of a random vertex) is updated in each step. More precisely, let $\mathcal{B} := \{B_1, \dots, B_K\}$ be a collection of subsets of vertices (called blocks) such that $V = \cup_{i=1}^K B_i$. Let α be a distribution over $\mathcal{B}$. The *(heat-bath) block dynamics* with respect to $(\mathcal{B},\alpha)$ is the Markov chain that, in each step, given a spin configuration σ_t, selects $B_i \in \mathcal{B}$ according to the distribution α and updates the configuration on B_i with a sample from the $\mu(\cdot \mid \sigma_t(V \setminus B_i))$; that is, from the conditional distribution on B_i given the spins of σ_t in $V \setminus B_i$. We denote this Markov chain (and its transition matrix) by $P_{\mathcal{B},\alpha}$. When the B_i's are each single vertices, and α is a uniform distribution over the blocks in $\mathcal{B}$, we obtain the Glauber dynamics. Our result for the mixing time of the block dynamics is the following.

Theorem 1.10. *Let $b > 0$, $\eta > 0$ and $\Delta \geq 3$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ. Let μ be a Gibbs distribution of a totally-connected spin system on G. Let $\mathcal{B} := \{B_1, \dots, B_K\}$ be any collection of blocks such that $V = \cup_{i=1}^K B_i$, and let α be a distribution over $\mathcal{B}$. If μ is η-spectrally independent and b-marginally bounded, then there exists a universal constant $C > 0$ such that the mixing time of block dynamics $P_{\mathcal{B},\alpha}$ satisfies:*

$$T_{mix}(P_{\mathcal{B},\alpha}) = O\Big(\alpha_{min}^{-1} \cdot \Big(\frac{C(\eta+1)^5 \Delta \log n}{b^6}\Big)^{3+\lceil \frac{2\eta}{b} \rceil}\Big),$$

where $\alpha_{min} = \min_{v \in V} \sum_{B \in \mathcal{B}} \alpha_B$.

See Theorem 5.2 for a more precise statement. Previous results for the block dynamics only apply to the bounded degree case [BCSV23, CP21, BCC+22], so Theorem 1.10 provides the first bounds for its mixing time in the unbounded degree setting.

2 Preliminaries

This section provides several definitions and background results we will refer to in our proofs.

2.1 Mixing times and modified log-Sobolev inequalities

Let P be an irreducible and aperiodic (i.e., ergodic) Markov chain with state space Ω and stationary distribution μ. Let us assume that P is reversible with respect to μ, and let

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \mu\|_{TV} := \max_{x \in \Omega} \max_{A \subseteq \Omega} |P^t(x, A) - \mu(A)|,$$

where $P^t(x, \cdot)$ denotes the distribution of the chain at time t assuming $x \in \Omega$ as the starting state; $\| \cdot \|_{TV}$ denotes the total variation distance. Note that with a slight abuse of notation we use P for both the Markov chain and its transition matrix. For $\varepsilon > 0$, let

$$T_{mix}(P, \varepsilon) := \min\{t > 0 : d(t) \leq \varepsilon\},$$

and the *mixing time of P* is defined as $T_{mix}(P) = T_{mix}(P, 1/4)$.

For functions $f, g : \Omega \to \mathbb{R}$, the *Dirichlet form* of a reversible Markov chain P with stationary distribution μ is defined as

$$\mathcal{E}_P(f, g) = \langle f, (I - P)g \rangle_\mu = \frac{1}{2} \sum_{x,y \in \Omega} \mu(x) P(x, y)(f(x) - f(y))(g(x) - g(y)),$$

where $\langle f, g \rangle_\mu := \sum_{x \in \Omega} f(x) g(x) \mu(x)$.

The spectrum of the ergodic and reversible Markov chain P is real, and we let $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1$ denote its eigenvalues. The (absolute) spectral gap of P is defined by $\mathrm{GAP}(P) = 1 - \max\{|\lambda_2|, |\lambda_{|\Omega|}|\}$. When P is positive semidefinite, we have

$$\mathrm{GAP}(P) = 1 - \lambda_2 = \inf \left\{ \frac{\mathcal{E}_P(f, f)}{\langle f, f \rangle_\mu} \mid f : \Omega \to \mathbb{R}, \langle f, f \rangle_\mu \neq 0 \right\}.$$

For P reversible and ergodic, we have the following standard comparison between the spectral gap and the mixing time

$$T_{mix}(P,\varepsilon) = \frac{1}{\mathrm{GAP}(P)} \cdot \log\big(\frac{1}{\varepsilon \mu_{min}}\big), \tag{2}$$

where $\mu_{min} := \min_{x\in\Omega} \mu(x)$.

The expected value of a function $f : \Omega \to \mathbb{R}_{\geq 0}$ with respect to μ is defined as $\mathrm{E}_\mu[f] = \sum_{x\in\Omega} f(x)\mu(x)$. Similarly, the entropy of the function with respect to μ is given by

$$\mathrm{Ent}_\mu(f) := \mathrm{E}_\mu\Big[f \log \frac{f}{\mathrm{E}_\mu[f]}\Big] = \mathrm{E}_\mu[f \log f] - \mathrm{E}_\mu[f \log(\mathrm{E}_\mu[f])].$$

We say that the Markov chain P satisfies a *modified log-Sobolev inequality* (MLSI) with constant ρ if for every function $f : \Omega \to \mathbb{R}_{\geq 0}$,

$$\rho \cdot \mathrm{Ent}_\mu(f) \leq \mathcal{E}_P(f, \log f).$$

The smallest ρ satisfying the inequality above is called the *modified log-Sobolev constant* of P and is denoted by $\rho(P)$. A well-known general relationship (see [DSC96, BT03]) shows that

$$\frac{1 - 2\mu_{min}}{\log(1/\mu_{min} - 1)} \mathrm{GAP}(P) \leq \rho(P) \leq 2\mathrm{GAP}(P). \tag{3}$$

For distributions μ and ν over Ω, the relative entropy of ν with respect to μ, denoted as $\mathcal{H}(\nu \mid \mu)$, is defined as $\mathcal{H}(\nu \mid \mu) := \sum_{x\in\Omega} \nu(x) \log \frac{\nu(x)}{\mu(x)}$. A Markov chain P with stationary distribution μ is said to satisfy discrete *relative entropy decay* with rate $r > 0$ if for all distributions ν:

$$\mathcal{H}(\nu P \mid \mu) \leq (1 - r)\mathcal{H}(\nu \mid \mu). \tag{4}$$

It is a standard fact (see, e.g., Lemma 2.4 in [BCP+22]) that when (4) holds, then $\rho(P) \geq r$, and

$$T_{mix}(P,\varepsilon) \leq \frac{1}{r} \cdot \Big(\log\log\big(\frac{1}{\mu_{min}}\big) + \log\big(\frac{1}{2\varepsilon}\big)\Big). \tag{5}$$

2.2 General spin system

We provide next a general definition for spin systems and introduce the notion of totally-connected systems.

Definition 2.1 (Spin system)**.** Let $G = (V, E)$ be a graph and $\mathcal{S} = \{1, \ldots, q\}$ a set of spins. Let $\Omega \subseteq \mathcal{S}^V$ be the set of possible spin configurations on G. We write σ_v for the spin assigned to v by σ. Given a configuration $\sigma \in \Omega$ and a subset Λ of V, we write $\sigma_\Lambda \in \mathcal{S}^\Lambda$ for the configuration of σ restricted to Λ. For a subset of vertices $\Lambda \subseteq V$, a *boundary condition* τ is an assignment of spins to (some) vertices in outer vertex boundary $\partial\Lambda \subseteq V \setminus \Lambda$ of Λ; namely, $\tau : (\partial\Lambda)_\tau \to \mathcal{S}$, with $(\partial\Lambda)_\tau \subseteq \partial\Lambda$. Note that a boundary condition is simply a pinning of a subset of vertices identified as being in the boundary of G. Given a boundary condition $\tau : (\partial V)_\tau \to \mathcal{S}$, the *Hamiltonian* $H : \Omega \to \mathbb{R}$ of a spin system is defined as

$$H(\sigma) = -\sum_{\{v,u\}\in E} K(\sigma_v, \sigma_u) - \sum_{\{v,u\}\in E: u\in V, v\in(\partial V)_\tau} K(\sigma_v, \tau_v) - \sum_{v\in V} U(\sigma_v), \tag{6}$$

where $K : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ and $U : \mathcal{S} \to \mathbb{R}$ are respectively the *symmetric edge interaction potential function* and the *spin potential function* of the system. The *Gibbs distribution* of a spin system with Hamiltonian H is defined as

$$\mu(\sigma) = \frac{1}{Z_H} e^{-H(\sigma)},$$

where $Z_H := \sum_{\sigma\in\Omega} e^{-H(\sigma)}$. We use Ω for the set of configurations σ satisfying $\mu(\sigma) > 0$.

The Potts model, as defined in the introduction, corresponds to the spin system with $q \geq 2$, $K(x, y) = \beta \cdot \mathbb{1}(x = y)$, and $U(\sigma_v) = 0$ for all $v \in V$. We focus on the ferromagnetic Ising model where $\beta > 0$ and $\mathcal{S} = \{-1, +1\}$. Another important spin system is the hardcore model that can be defined by setting $\mathcal{S} = \{1, 0\}$, $K(x, y) = \infty$ if $x = y = 1$ and $K(x, y) = 0$ otherwise, and $U(x) = \mathbb{1}(x = 1) \cdot \ln \lambda$, where $\lambda > 0$ is referred to as the *fugacity* parameter of the model.

We restrict attention to *totally-connected* spin systems, as this ensures that the Glauber dynamics, the systematic scan dynamics, and the block dynamics are all irreducible Markov chains (and thus ergodic).

Definition 2.2. For a subset C_U of partial configurations on $U \subseteq V$, let $H[C_U] = (C_U, E[C_U])$ be the induced subgraph where $E[C_U]$ consists of all pairs of configurations on C_U that differ at exactly one vertex. We say that C_U is connected when $H[C_U]$ is connected. For a pinning τ on $\Lambda \subseteq V$, we say $\Omega^\tau_{V \setminus \Lambda}$ is connected if $H[\Omega^\tau_{V \setminus \Lambda}]$ is connected. A distribution μ over $\mathcal{S}^V$ is *totally-connected* if for every $\Lambda \subseteq V$ and every pinning τ on Λ, $\Omega^\tau_{V \setminus \Lambda}$ is connected.

3 Swendsen-Wang dynamics on general graphs

In this section, we consider the SW dynamics for the q-state ferromagnetic Potts models on general graphs. In particular, we establish Theorem 1.2 from the introduction, which is a direct corollary of the following more general result.

Theorem 3.1. *Let $q \geq 2$, $\beta > 0$, $\eta > 0$, $b > 0$, $\Delta \geq 3$, and $\chi \geq 2$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ and chromatic number χ. Let μ_{Potts} be the Gibbs distribution of the q-state ferromagnetic Potts model on G with parameter β. If μ_{Potts} is η-spectrally independent and b-marginally bounded, then there exists a universal constant $C > 1$ such that the modified log-Sobolev constant of the SW dynamics satisfies:*

$$\rho(P_{SW}) = \Omega\left(\frac{b^{2+6\kappa}}{\chi \cdot (C\Delta \log n)^\kappa \cdot (\eta + 1)^{5\kappa}}\right),$$

where $\kappa = 2 + \lceil \frac{2\eta}{b} \rceil$, and

$$T_{mix}(P_{SW}) = O\big(\chi \cdot (C\Delta \log n)^\kappa \cdot (\eta + 1)^{5\kappa} b^{-2-6\kappa} \cdot \log n\big).$$

Theorem 1.2 follows from this theorem by noting that $\chi \leq \Delta$ and that under the assumptions $\eta = O(1)$ and $\beta\Delta = O(1)$, we have $b = O(1)$ and $\kappa = O(1)$.

Remark 1. When Δ is small, i.e., $\Delta = o(\log n)$, we can obtain slightly better bounds on $\rho(P_{SW})$ and $T_{mix}(P_{SW})$ and replace the $(C\Delta \log n)^\kappa$ factor by a factor of $(C\Delta)^{8+4\lceil \frac{2\eta}{b} \rceil}$.

Before proving Theorem 3.1, we provide a number of definitions and required background results in Section 3.1. We then give the proof of Theorem 3.1 in Sections 3.2, 3.3, and 3.4, and include some applications of this result in Section 3.5.

3.1 Factorization of entropy

We present next several factorizations of the entropy functional $\text{Ent}_\mu(f)$, which are instrumental in establishing the decay of the relative entropy for the SW dynamics. We introduce some useful notations first. For a pinning τ in $V \setminus \Lambda$ (i.e., $\tau \in \Omega_{V \setminus \Lambda}$), we let $\mu^\tau_\Lambda(\cdot) := \mu(\cdot \mid \sigma_{V \setminus \Lambda} = \tau)$. Given a function $f : \Omega \to \mathbb{R}_{\geq 0}$, subsets of vertices $B \subseteq \Lambda \subset V$, and $\tau \in \Omega_{V \setminus \Lambda}$, the function $f^\tau_B : \Omega^\tau_B \to \mathbb{R}_{\geq 0}$ is defined by:

$$f^\tau_B(\sigma) = \text{E}_{\xi \sim \mu^\tau_{\Lambda \setminus B}}[f(\tau \cup \xi \cup \sigma)].$$

If $B = \Lambda$, we often write f^τ for f_B^τ, and if $\tau = \emptyset$, then we use f_B for f_B^τ. We use $\mathrm{Ent}_B^\tau(f^\tau)$ to denote $\mathrm{Ent}_{\mu_B^\tau}(f^\tau)$, and if the pinning τ on $V \setminus B$ is from a distribution π over $\Omega_{V\setminus B}$, we use $\mathrm{E}_{\tau\sim\pi}[\mathrm{Ent}_B^\tau(f^\tau)]$ to denote the expected value of the function f on S over the random pinning τ.

Various forms of entropy factorization arise from bounding $\mathrm{Ent}_\mu(f)$ by different (weighted) sums of restricted entropies of the function f. The first one we introduced, is the so-called *ℓ-uniform block factorization of entropy* of *ℓ-UBF*. For an integer $\ell \le n$, ℓ-UBF holds for μ with constant C_{UBF} if for all functions $f : \Omega \to \mathbb{R}_{\ge 0}$,

$$\frac{\ell}{n} \cdot \mathrm{Ent}_\mu(f) \le C_{\mathrm{UBF}} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S\in\binom{V}{\ell}} \mathrm{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Ent}_S^\tau(f^\tau)\right], \tag{7}$$

where $\binom{V}{\ell}$ denotes the collection of all subsets of V of size ℓ. An important special case is when $\ell = 1$, in which case (7) is called *approximate tensorization of entropy (AT)*; this special case has been quite useful for establishing optimal mixing time bounds for the Glauber dynamics in various settings (see, e.g., [Mar19, CMT14, Ces01, Mar99]). In recent works, a key step for obtaining AT has been to first establish ℓ-UBF for some large ℓ. The following result will be useful for us.

Theorem 3.2 ([CLV21], [BCC+22])**.** *Let b and η be fixed. For $\theta \in (0,1)$ and $n \ge \frac{2}{\theta}(\frac{4\eta}{b^2}+1)$, the following holds. If the Gibbs distribution μ of a totally-connected spin system on an n-vertex graph is η-spectrally independent and b-marginally bounded, then $\lceil\theta n\rceil$-UBF holds with $C_{\mathrm{UBF}} = (e/\theta)^{\lceil\frac{2\eta}{b}\rceil}$.*

Another useful notion is the *k-partite factorization of entropy* or KPF. Let $U_1,\dots,U_k$ be k disjoint independent sets of V such that $\bigcup_{i=1}^k U_i = V$. We say μ satisfies KPF with constant C_{KPF} if for all functions $f : \Omega \to \mathbb{R}_{\ge 0}$,

$$\mathrm{Ent}_\mu(f) \le C_{\mathrm{KPF}} \sum_{i=1}^k \mathrm{E}_{\tau\sim\mu_{V\setminus U_i}}\left[\mathrm{Ent}_{U_i}^\tau(f^\tau)\right].$$

KPF was introduced in [BCC+22], where it was used to analyze global Markov chains. The interplay between KPF and UBF is intriguing and is further explored in this paper.

3.2 Proof of main result for the SW dynamics: Theorem 3.1

The main technical contribution in the proof of Theorem 3.1 is establishing KPF with a better (i.e., smaller) constant C_{KPF}. As in [BCC+22], KPF is then used to derive an improved "edge-spin" factorization of entropy which is known to imply the desired bounds on the modified log-Sobolev constant and on the mixing time of the SW dynamics.

Theorem 3.3. *For a totally-connected and b-marginally bounded Gibbs distribution μ that satisfies η-spectral independence on an n-vertex graph $G = (V,E)$ of maximum degree $\Delta \ge 3$, if b and η are constants independent of Δ and n, then there exists a constant $c = c(\eta, b) > 0$ such that k-partite factorization of entropy holds for μ with constant $C_{\mathrm{KPF}} = (\Delta \log n)^c$. Specifically, for a set of k disjoint independent sets $V_1,\dots,V_k$ such that $\bigcup_{j=1}^k V_j = V$, we have*

$$\mathrm{Ent}_\mu(f) \le \left(\frac{C(\eta+1)^5\Delta\log n}{b^6}\right)^\kappa \cdot \sum_{j=1}^k \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}[\mathrm{Ent}_{V_j}^\tau(f^\tau)], \text{ and} \tag{8}$$

$$\mathrm{Ent}_\mu(f) \le \left(\frac{C(\eta+1)^5\Delta^4}{b^6}\right)^\kappa \cdot \sum_{j=1}^k \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}[\mathrm{Ent}_{V_j}^\tau(f^\tau)], \tag{9}$$

where $\kappa := 2 + \lceil\frac{2\eta}{b}\rceil$ and $C > 0$ is a universal constant.

Remark 2. Let $\mathcal{B} = \{B_1, \dots, B_k\}$ be a collection of disjoint independent sets such that $V = \bigcup_{i=1}^{k} B_i$. The independent set dynamics $P_{\mathcal{B}}$ is a heat-bath block dynamics w.r.t. $\mathcal{B}$ and a uniform distribution over $\mathcal{B}$. If μ satisfies k-partite factorization of entropy with C_{KPF}, then $P_{\mathcal{B}}$ satisfies a relative entropy decay with rate $r \geq 1/(k \cdot C_{\mathrm{KPF}})$. See Lemma 5.1 for the more general statement.

As mentioned, KPF was first studied in [BCC+22]; the constant proved there was

$$C_{\mathrm{KPF}} = b^{-O(\Delta)} \cdot (\Delta/b)^{O(\eta/b)},$$

so our new bound improves the dependence on Δ from exponential to polynomial. The proof of Theorem 3.3 is given in two parts. In Section 3.3, we prove (8), whereas (9) is proved in Appendix A.

With KPF on hand, the next step in the proof of Theorem 3.1 relies on the so-called edge-spin factorization of entropy. Let $\Omega_J := \Omega \times \{0,1\}^E$ be the set of joint configurations (σ, A) corresponding to pairs of a spin configuration $\sigma \in \Omega$ and an *edge configuration* (a subset of edges in a graph) $A \subseteq E$. For a q-state Potts model μ_{Potts} with parameter $p = 1 - e^{-\beta}$, we use ν to denote the *Edwards-Sokal* measure on Ω_J given by

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

where $\sigma \sim A$ is the event that every edge in A has its two endpoints with the same spin in σ, and $Z_J := \sum_{(A,\sigma)\in\Omega_J}(1-p)^{|E|-|A|}p^{|A|}\mathbf{1}(\sigma \sim A)$ is a normalizing constant. Let $\nu(\cdot \mid \sigma)$ and $\nu(\cdot \mid A)$ denote the conditional measures obtained from ν by fixing the spin configuration to be σ or fixing the edge configuration to be A respectively. For a function $f : \Omega_J \to \mathbb{R}_{\geq 0}$, let $f^\sigma : \{0,1\}^{|E|} \to \mathbb{R}_{\geq 0}$ be the function given by $f^\sigma(A) = f(\sigma \cup A)$, and let $f^A : \Omega \to \mathbb{R}_{\geq 0}$ be the function given by $f^A(\sigma) = f(\sigma \cup A)$. We say that *edge-spin factorization of entropy* holds with constant C_{ES} if for all functions $f : \Omega_J \to \mathbb{R}_{\geq 0}$,

$$\mathrm{Ent}_\nu(f) \leq C_{\mathrm{ES}} \left(\mathrm{E}_{(\sigma,A)\sim\nu} \left[\mathrm{Ent}_{A\sim\nu(\cdot|\sigma)}(f^\sigma) \right] + \mathrm{E}_{(\sigma,A)\sim\nu} \left[\mathrm{Ent}_{\sigma\sim\nu(\cdot|A)}(f^A) \right] \right). \tag{10}$$

The following result from [BCC+22] will be useful for us.

Lemma 3.4 (Theorem 6.1 [BCC+22])**.** *Suppose the q-state ferromagnetic Potts model with parameter β on a graph G of maximum degree is $\Delta \geq 3$ satisfies KPF with constant C_{KPF}. Then, the edge-spin factorization of entropy holds with constant $C_{\mathrm{ES}} = O(\beta\Delta k e^{\beta\Delta}) \cdot C_{\mathrm{KPF}}$.*

Remark 3. The original bound for C_{ES} stated in [BCC+22] is actually $O(\beta\Delta^2 e^{\beta\Delta}) \cdot C_{\mathrm{KPF}}$, but in the proof there, one factor k is replaced with Δ as its upper bound. Since we do not assume Δ to be a constant, we avoid such an upper bound. We also remark that the exponential dependence of C_{ES} on $\beta\Delta$ can probably be improved, but in our applications $\beta\Delta = O(1)$, so this would not represent a tangible improvement.

The final ingredient in the proof of Theorem 3.1 is the following.

Lemma 3.5 (Lemma 1.8 [BCP+22])**.** *Suppose edge-spin factorization of entropy holds with constant C_{ES}. Then, the SW dynamics P_{SW} satisfies the relative entropy decay with rate $\Omega\left(\frac{1}{C_{\mathrm{ES}}}\right)$.*

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. By Theorem 3.3, μ_{Potts} satisfies χ-partite factorization of entropy with constant

$$C_{\mathrm{KPF}} = \left(\frac{C(\eta+1)^5\Delta\log n}{b^6}\right)^\kappa,$$

where $C > 0$ is a universal constant. It follows from Lemma 3.4 and Lemma 3.5 that the SW dynamics satisfies (4) with

$$r = \Omega\left(\frac{b^{6\kappa}}{\chi\beta\Delta e^{\beta\Delta} \cdot C^\kappa(\eta+1)^{5\kappa} \cdot (\Delta\log n)^\kappa}\right).$$

Note that $b \leq q^{-1}e^{-\beta\Delta}$, and so $\beta\Delta e^{\beta\Delta} \leq e^{2\beta\Delta} \leq b^{-2}$. Therefore, we obtain the desired bound for MLSI constant, and the mixing time bound follows from (5). $\square$

3.3 Proof of the main technical theorem: Theorem 3.3

Recall that given a function $f : \Omega \to \mathbb{R}_{\geq 0}$, subsets of vertices $B \subseteq \Lambda \subset V$, and $\tau \in \Omega_{V\setminus\Lambda}$, the function $f_B^\tau : \Omega_B^\tau \to \mathbb{R}_{\geq 0}$ is defined by

$$f_B^\tau(\sigma) = \mathrm{E}_{\xi \sim \mu^\tau_{\Lambda\setminus B}}[f(\tau \cup \xi \cup \sigma)].$$

In the proof of Theorem 3.3 we use several facts, which we compile next.

Let $S \subseteq V$ be a subset of vertices. Let $S_1, \dots, S_m \subseteq V$ denote the connected components of S. For a vertex $v \in V$, let $C_S(v)$ the unique connected component S_i that contains v, if such component exists, otherwise set $C_S(v)$ to be the empty set. When S is chosen uniformly at random among all subsets of size $\lceil \theta n \rceil$, the following exponential tail bound for $|C_S(v)|$ was established in [CLV21].

Lemma 3.6 (Lemma 4.3, [CLV21])**.** *Let $G = (V, E)$ be an n-vertex graph of maximum degree at most Δ. Then for any $v \in V$ and every integer $k \geq 0$ we have*

$$\Pr_S[|C_S(v)| = k] \leq \frac{\ell}{n} \cdot (2e\Delta\theta)^{k-1},$$

where the probability $\Pr_S[\cdot]$ is taken over a uniformly random subset $S \subseteq V$ of size $\ell = \lceil \theta n \rceil$.

Lemma 3.7. *Let μ be a totally-connected and b-marginally bounded distribution over $[q]^n$. If μ is η-spectrally independent, then the Glauber dynamics for μ has spectral gap at least*

$$\left(\frac{2b^4}{(\lceil 2\eta \rceil + 2)^4} \cdot \frac{1}{n}\right)^{1+\lceil 2\eta \rceil}. \tag{11}$$

Remark 4. Lemma 3.7 is similar to Theorem 1.3 in [ALOG20] (for 2-spin systems), Theorem 6 in [CGSV21] (for colorings), and Theorem 3.2 in [FGYZ22] (for a different notion of spectral independence). For completeness, we provide a proof in Appendix B.

Lemma 3.8. *Let μ be a b-marginally bounded distribution over $[q]^n$. If the Glauber dynamics for μ has spectral gap γ, then μ satisfies KPF with constant*

$$C_{KPF} \leq \frac{3n \log(b^{-1})}{\gamma}. \tag{12}$$

The proof of Lemma 3.8 is standard and is provided in Appendix B. We proceed to prove (8) from Theorem 3.3. With a slightly different argument, we will establish (9) in Appendix A, which is a better upper bound only when $\Delta = o(\log n)$.

Proof of (8) *in Theorem 3.3.* It follows from Lemma 3.7 and Lemma 3.8 that

$$C_{KPF} \leq \frac{3(\lceil 2\eta \rceil + 2)^{4(1+\lceil 2\eta \rceil)}}{(2b^4)^{2+\lceil 2\eta \rceil}} \cdot n^{2+\lceil 2\eta \rceil}.$$

If $\Delta > \frac{b^2 n}{10e(4\eta + b^2)}$, letting $\kappa := 2 + \lceil \frac{2\eta}{b} \rceil$, then we establish the theorem since

$$\frac{3(\lceil 2\eta \rceil + 2)^{4(1+\lceil 2\eta \rceil)}}{(2b^4)^{2+\lceil 2\eta \rceil}} \cdot n^{2+\lceil 2\eta \rceil} \leq \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^\kappa} \cdot \left(\frac{10e(4\eta + b^2)}{b^2}\right)^\kappa \cdot \Delta^\kappa \leq \frac{(240e)^{4\kappa} \cdot (\lceil \eta \rceil + 1)^{5\kappa} \cdot \Delta^\kappa}{b^{6\kappa}}.$$

Thus, we assume $\Delta \leq \frac{b^2 n}{10e(4\eta + b^2)}$. Let $V_1, \dots, V_k \subseteq V$ be disjoint independent sets such that $\bigcup_j V_j = V$. We take $\theta = \frac{1}{5e\Delta}$ so that $\frac{2}{n} \cdot (\frac{4\eta}{b^2} + 1) < \theta$. Let S be a subset of vertices of size $\lceil \theta n \rceil$ chosen uniformly at random from all the subsets of size $\lceil \theta n \rceil$. Let $S_1, \ldots, S_m \subseteq V$ be the connected components of S. Theorem 3.2 implies that $\lceil \theta n \rceil$-UBF holds with constant

$$C_{\mathrm{UBF}} = \left(\frac{e}{\theta}\right)^{\lceil \frac{2\eta}{b} \rceil} = \left(5e^2\Delta\right)^{\lceil \frac{2\eta}{b} \rceil}, \tag{13}$$

and so for any function $f : \Omega \to \mathbb{R}_{\geq 0}$ we have

$$\mathrm{Ent}_\mu(f) \leq \left(5e^2\Delta\right)^{1+\lceil \frac{2\eta}{b} \rceil} \mathrm{E}_S\left[\mathrm{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Ent}_S^\tau(f^\tau)\right]\right], \tag{14}$$

where E_S denotes the expectation over the random subset S. To bound the right-hand side of (14), we use the following fact, which we prove later in Section 3.4.

Lemma 3.9. *Let $V_1, \ldots, V_k$ be disjoint independent sets such that $\bigcup_{j=1}^k V_j = V$. Let $S \subseteq V$ be a subset of vertices. Let $S_1, \ldots, S_m \subseteq S$ be the connected components of the subgraph induced by S. Suppose that for $S_i \subseteq S$, $\Gamma(S_i)$ takes the minimum value such that the following inequality holds for an arbitrary pinning $\tau \in \Omega_{V\setminus S_i}$ and any function $g : \Omega_{S_i}^\tau \to \mathbb{R}_{\geq 0}$:*

$$\mathrm{Ent}_{S_i}^\tau(g) \leq \Gamma(S_i) \sum_{j=1}^k \mathrm{E}_{\xi\sim\mu_{S_i\setminus V_j}^\tau}\left[\mathrm{Ent}_{V_j\cap S_i}^{\xi\cup\tau}(g_{S_i\cap V_j}^\xi)\right]. \tag{15}$$

Then for any function $f : \Omega \to \mathbb{R}_{\geq 0}$,

$$\mathrm{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Ent}_S^\tau(f^\tau)\right] \leq \sum_{j=1}^k \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}\left[\mathrm{Ent}_{V_j}^\tau(f^\tau)\right] \cdot \max_{S_i\subseteq S} \Gamma(S_i). \tag{16}$$

From (14) and Lemma 3.9, we have

$$\mathrm{Ent}_\mu(f) \leq \left(5e^2\Delta\right)^\kappa \sum_{j=1}^k \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}\left[\mathrm{Ent}_{V_j}^\tau(f^\tau)\right] \cdot \mathrm{E}_S\left[\max_{S_i\subseteq S} \Gamma(S_i)\right]. \tag{17}$$

To show (8), it remains to provide an upper bound for $\mathrm{E}_S\left[\max_{S_i\subseteq S}\Gamma(S_i)\right]$.

By assumption, μ is η-spectrally independent and b-marginally bounded. These properties, by definition, are preserved under any pinning. In particular, for any $S_i \subseteq S$ and an arbitrary pinning $\tau \in \Omega_{V\setminus S_i}$, $\mu_{S_i}^\tau$ is still η-spectrally independent and b-marginally bounded. Hence, by Lemma 3.7 and Lemma 3.8, we have

$$\Gamma(S_i) \leq \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^\kappa} \cdot |S_i|^\kappa,$$

and

$$\mathrm{E}_S\left[\max_{S_i\subseteq S}\Gamma(S_i)\right] \leq b_1 \mathrm{E}_S\left[\max_{S_i\subseteq S}|S_i|^\kappa\right] = b_1 \mathrm{E}_S\left[\max_{v\in S}|C_S(v)|^\kappa\right], \tag{18}$$

where $b_1 := \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^\kappa}$. To estimate the expectation on the right-hand side of (18), we first expand the expectation and apply a union bound as follows:

$$\begin{aligned}
\mathrm{E}_S\left[\max_{v\in S}|C_S(v)|^\kappa\right] &= \sum_{x=0}^{|S|} x^\kappa \cdot \mathrm{Pr}_S\left[\max_{v\in S}|C_S(v)| = x\right] \\
&\leq (2\log_2|S|)^\kappa + \sum_{x=2\log_2|S|}^{|S|} x^\kappa \cdot \mathrm{Pr}_S\left[\max_{v\in S}|C_S(v)| = x\right] \\
&\leq (2\log_2|S|)^\kappa + \sum_{x=2\log_2|S|}^{|S|} x^\kappa \cdot \sum_{v\in S}\mathrm{Pr}_S\left[|C_S(v)| = x\right].
\end{aligned} \tag{19}$$

Then, applying Lemma 3.6 and noting that $\theta < 1/(4e\Delta)$, we obtain

$$
\begin{aligned}
\sum_{x=2\log_2|S|}^{|S|} x^\kappa \cdot \sum_{v\in S} \Pr_S\left[|C_S(v)| = x\right] &\le \lceil \theta n \rceil \sum_{x=2\log_2|S|}^{|S|} x^\kappa (2e\Delta\theta)^{x-1} \\
&= \frac{\lceil \theta n \rceil}{2e\Delta\theta} \cdot (2e\Delta\theta)^{2\log_2|S|} \sum_{x=2\log_2|S|}^{|S|} x^\kappa (2e\Delta\theta)^{x-2\log_2|S|} \\
&\le \frac{1}{2|S|e\Delta} \sum_{x=0}^{|S|-2\log_2|S|} (x+2\log_2|S|)^\kappa 2^{-x} \\
&\le \frac{1}{2|S|e\Delta}\left[\sum_{x=0}^{\log_2|S|-1} (x+2\log_2|S|)^\kappa + \sum_{x=\log_2|S|}^{|S|-2\log_2|S|} \frac{(x+2\log_2|S|)^\kappa}{|S|\cdot 2^{x-\log_2|S|}}\right] \\
&\le \frac{1}{2|S|e\Delta}\left[(3\log_2|S|)^\kappa + \sum_{x=0}^{|S|-3\log_2|S|} \frac{(x+3\log_2|S|)^\kappa}{|S|\cdot 2^x}\right]. \qquad (20)
\end{aligned}
$$

When $|S| = \omega(1)$, $(3\log_2|S|)^{1+\kappa}/|S| < 1$. Also, for any integer $x \ge 0$, $\frac{(x+3\log_2|S|)^\kappa}{|S|\cdot 2^x} < 1$, so the last sum in (20) is less than $|S|$. Therefore, by (18), (19) and (20) we have

$$
\mathrm{E}_S\left[\max_{S_i\subseteq S} \Gamma(S_i)\right] \le b_1 \cdot \left[(2\log_2|S|)^\kappa + 1\right] \le b_1(3\log_2|S|)^\kappa. \qquad (21)
$$

These bounds together with (17) imply that

$$
C_{\mathrm{KPF}} \le b_1(3\log_2 n)^\kappa \cdot \left(5e^2\Delta\right)^\kappa = 3\cdot\left(\frac{15e^2}{2}\right)^\kappa \cdot \frac{(\lceil 2\eta\rceil + 2)^{4\kappa}}{b^{4\kappa}} \cdot (\Delta\log_2 n)^\kappa,
$$

establishing the desired bound. When $|S| = O(1)$, the left-hand side of (21) can be bounded by an absolute constant, and the result follows from (17). ◻

3.4 Entropy factorization: Proof of Lemma 3.9

We proceed with the proof of Lemma 3.9 by first presenting several facts that will be useful.

Lemma 3.10 (Lemma 2.7, [BCC+22]). *Let $\Lambda = A\cup B \subseteq V$, $\tau \in \Omega_{V\setminus\Lambda}$, and assume μ_Λ^τ is a product measure $\mu_\Lambda^\tau = \mu_A^\tau \otimes \mu_B^\tau$. For all $U \subset B$ and any $f : \Omega \to \mathbb{R}_{\ge 0}$,*

1. $\mathrm{Ent}_A^\tau(f_A^\tau) = \mathrm{E}_{\gamma\sim\mu_B^\tau}\left[\mathrm{Ent}_A^{\gamma\cup\tau}(f_A^\tau)\right]$.
2. $\mathrm{E}_{\gamma\sim\mu_B^\tau}\left[\mathrm{Ent}_A^{\gamma\cup\tau}(f_A^\tau)\right] \le \mathrm{E}_{\gamma\sim\mu_U^\tau}\left[\mathrm{Ent}_A^{\gamma\cup\tau}(f_A^{\gamma\cup\tau})\right]$.

Lemma 3.11 (Lemma 3.1, [CP21]). *Let $\Lambda_0 = \emptyset$. For any $\Lambda_1 \subset \dots \Lambda_m \subset \Lambda \subseteq V$, any $\tau \in \Omega_{V\setminus\Lambda}$ and any $f : \Omega_\Lambda^\tau \to \mathbb{R}_{\ge 0}$,*

$$
\sum_{i=1}^m \mathrm{E}_{\gamma\sim\mu_{\Lambda\setminus\Lambda_i}^\tau}\left[\mathrm{Ent}_{\Lambda_i\setminus\Lambda_{i-1}}^{\tau\cup\gamma}(f_{\Lambda_i\setminus\Lambda_{i-1}}^\gamma)\right] = \mathrm{E}_{\gamma\sim\mu_{\Lambda\setminus\Lambda_m}^\tau}\left[\mathrm{Ent}_{\Lambda_m}^{\tau\cup\gamma}(f^\gamma)\right].
$$

The following corollary directly follows from this fact, by taking $\Lambda_1 = A, \Lambda_2 = B$ and $m = 2$.

Corollary 3.12. *Let A, B and Λ be subsets of vertices such that $A \subset B \subset \Lambda \subseteq V$. For any $\tau \in \Omega_{V\setminus\Lambda}$ and any $f : \Omega^\tau_\Lambda \to \mathbb{R}_{\geq 0}$,*

$$\mathrm{E}_{\gamma\sim\mu^\tau_{\Lambda\setminus A}}\left[\mathrm{Ent}^{\gamma\cup\tau}_A(f^\gamma)\right] \leq \mathrm{E}_{\gamma\sim\mu^\tau_{\Lambda\setminus B}}\left[\mathrm{Ent}^{\gamma\cup\tau}_B(f^\gamma)\right].$$

We are now ready to prove Lemma 3.9.

Proof of Lemma 3.9. Note that $\mu^\tau_S = \otimes^m_{i=1}\mu^\tau_{S_i}$ is a product measure. For $i \geq 1$, let $S_{\leq i} := S_1 \cup \cdots \cup S_i$. For $i > 1$, we let $S_{<i} := S_1 \cup \cdots \cup S_{i-1}$, and we set $S_{<1} := \emptyset$ for convenience. As a direct consequence of applying Lemma 3.11 and applying Lemma 3.10(1), we have the following identity for any $f : \Omega \to \mathbb{R}_{\geq 0}$:

$$\mathrm{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Ent}^\tau_S(f^\tau)\right] = \sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{V\setminus(S_{\leq i})}}\left[\mathrm{Ent}^\tau_{S_i}(f^\tau_{S_i})\right] = \sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{V\setminus(S_{\leq i})}}\left[\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}\left[\mathrm{Ent}^{\tau\cup\gamma}_{S_i}(f^\tau_{S_i})\right]\right]. \tag{22}$$

On the other hand, setting $g = f^\tau_{S_i}$ in (15), then for any $\gamma \in \Omega^\tau_{S_{<i}}$ we obtain that

$$\mathrm{Ent}^{\tau\cup\gamma}_{S_i}(f^\tau_{S_i}) \leq \Gamma(S_i)\sum^k_{j=1}\mathrm{E}_{\xi\sim\mu^{\gamma\cup\tau}_{S_i\setminus V_j}}\left[\mathrm{Ent}^{\xi\cup\tau\cup\gamma}_{V_j\cap S_i}(f^{\tau\cup\xi}_{S_i\cap V_j})\right]. \tag{23}$$

Combining (22) and (23) yields

$$\begin{aligned}\mathrm{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Ent}^\tau_S(f^\tau)\right] &\leq \sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{V\setminus S_{\leq i}}}\left[\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}\left[\Gamma(S_i)\sum^k_{j=1}\mathrm{E}_{\xi\sim\mu^{\tau\cup\gamma}_{S_i\setminus V_j}}\left[\mathrm{Ent}^{\xi\cup\gamma\cup\tau}_{V_j\cap S_i}(f^{\xi\cup\tau}_{S_i\cap V_j})\right]\right]\right]\\ &= \sum^k_{j=1}\sum^m_{i=1}\Gamma(S_i)\mathrm{E}_{\tau\sim\mu_{V\setminus S_{\leq i}}}\mathrm{E}_{\xi\sim\mu^\tau_{S_i\setminus V_j}}\mathrm{E}_{\gamma\sim\mu^{\tau\cup\xi}_{S_{<i}}}\left[\mathrm{Ent}^{\xi\cup\gamma\cup\tau}_{V_j\cap S_i}(f^{\xi\cup\tau}_{S_i\cap V_j})\right]\\ &\leq \sum^k_{j=1}\max_i\Gamma(S_i)\sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{(V\setminus S_{\leq i})\cup(S_i\setminus V_j)}}\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}\left[\mathrm{Ent}^{\gamma\cup\tau}_{V_j\cap S_i}(f^\tau_{S_i\cap V_j})\right].\end{aligned} \tag{24}$$

We show next that for any $j = 1, \ldots, k$, the following inequality holds:

$$\sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{(V\setminus S_{\leq i})\cup(S_i\setminus V_j)}}\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}\left[\mathrm{Ent}^{\gamma\cup\tau}_{V_j\cap S_i}(f^\tau_{S_i\cap V_j})\right] \leq \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}\left[\mathrm{Ent}^\tau_{V_j}(f^\tau)\right]. \tag{25}$$

Given a pinning $\tau \sim \mu_{V\setminus((S_i\cap V_j)\cup S_{<i})}$, $\mu_{S_i\cap V_j}$ and $\mu_{S_{<i}}$ are independent. By applying Lemma 3.10(2) to $\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}[\mathrm{Ent}^{\gamma\cup\tau}_{V_j\cap S_i}(f^\tau_{S_i\cap V_j})]$, we have

$$\sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{(V\setminus S_{\leq i})\cup(S_i\setminus V_j)}}\mathrm{E}_{\gamma\sim\mu^\tau_{S_{<i}}}\left[\mathrm{Ent}^{\gamma\cup\tau}_{V_j\cap S_i}(f^\tau_{S_i\cap V_j})\right] \leq \sum^m_{i=1}\mathrm{E}_{\tau\sim\mu_{(V\setminus S_{\leq i})\cup(S_i\setminus V_j)}}\mathrm{E}_{\xi\sim\mu^\tau_{(S_{<i})\setminus V_j}}\left[\mathrm{Ent}^{\tau\cup\xi}_{S_i\cap V_j}(f^{\tau\cup\xi}_{S_i\cap V_j})\right]. \tag{26}$$

Letting $\phi = \tau\cup\xi$, and by applying Lemma 3.10(1) to $\mathrm{Ent}^\phi_{S_i\cap V_j}(f^\phi_{S_i\cap V_j})$ we also have

$$\sum^m_{i=1}\mathrm{E}_{\phi\sim\mu_{V\setminus((S_{\leq i})\cap V_j)}}\left[\mathrm{Ent}^\phi_{S_i\cap V_j}(f^\phi_{S_i\cap V_j})\right] = \sum^m_{i=1}\mathrm{E}_{\phi\sim\mu_{V\setminus(S_{\leq i})\cap V_j}}\mathrm{E}_{\psi\sim\mu^\phi_{(S_{<i})\cap V_j}}\left[\mathrm{Ent}^{\phi\cup\psi}_{S_i\cap V_j}(f^\phi_{S_i\cap V_j})\right]. \tag{27}$$

Also, the following identity follows from Lemma 3.10(1) and Lemma 3.11 as in the way of obtaining (22):

$$\mathrm{E}_{\tau\sim\mu_{V\setminus(S\cap V_j)}}\left[\mathrm{Ent}^\tau_{S\cap V_j}(f^\tau)\right] = \sum^m_{i=1}\mathrm{E}_{\phi\sim\mu_{V\setminus(S_{\leq i}\cap V_j)}}\mathrm{E}_{\psi\sim\mu^\phi_{(S_{<i})\cap V_j}}\left[\mathrm{Ent}^{\phi\cup\psi}_{S_i\cap V_j}(f^\phi_{S_i\cap V_j})\right]. \tag{28}$$

Finally, it follows from Corollary 3.12 that

$$\mathrm{E}_{\tau\sim\mu_{V\setminus(S\cap V_j)}}\left[\mathrm{Ent}^\tau_{S\cap V_j}(f^\tau)\right] \leq \mathrm{E}_{\tau\sim\mu_{V\setminus V_j}}\left[\mathrm{Ent}^\tau_{V_j}(f^\tau)\right], \tag{29}$$

so (25) follows from (26), (27), (28) and (29). Therefore, we obtain (16) by (24) and (25). $\square$

3.5 Applications of Theorem 3.1

In this section, we prove Corollary 1.3 from the introduction and present another application of Theorem 3.1 concerning the SW dynamics on a random graph generated from the classical Erdős-Rényi $G(n, p)$ model. For this, we first define Dobrushin's influence matrix.

Definition 3.13. The *Dobrushin influence matrix* $A \in \mathbb{R}^{n\times n}$ is defined by $A(u, u) = 0$ and for $u \neq v$,

$$A(u, v) = \max_{(\sigma,\tau)\in S_{u,v}} d_{TV}(\mu_v(\cdot \mid \sigma), \mu_v(\cdot \mid \tau)),$$

where $S_{u,v}$ contains the set of all pairs of partial configurations (σ, τ) in $\Omega_{V\setminus\{v\}}$ that can only disagree at u, namely, $\sigma_w = \tau_w$ if $w \neq u$.

It is known that an upper bound on the spectral norm of A implies spectral independence. In particular, we have the following result from [BCC+22].

Proposition 3.14 (Theorem 1.13, [BCC+22])**.** *If the Dobrushin influence matrix A of a distribution μ satisfies $\|A\| \leq 1 - \varepsilon$ for some $\varepsilon > 0$, then μ is spectral independent with constant $\eta = 2/\varepsilon$.*

For the ferromagnetic Ising model, $\beta_u(\Delta) := \ln \frac{\Delta}{\Delta-2}$ corresponds to the threshold value of the parameter β for the uniqueness/non-uniqueness phase transition on the Δ-regular tree. For the anti-ferromagnetic Ising model, the phase transition occurs at $\bar{\beta}_u(\Delta) := -\ln \frac{\Delta}{\Delta-2}$. If $\bar{\beta}_u(\Delta)(1-\delta) < \beta < \beta_u(\Delta)(1-\delta)$, we say the Ising model satisfies the δ*-uniqueness condition.* On a bounded degree graph, $\|A\| \leq 1-\delta$ for the Ising model is a strictly stronger condition than δ-uniqueness condition. However, due to the observation made in [AJK+22], if $\Delta \to \infty$, the two conditions are roughly equivalent.

Proposition 3.15. *The Ising model with parameter $\bar{\beta}_u(\Delta)(1-\delta) < \beta < \beta_u(\Delta)(1-\delta)$ and $\Delta \to \infty$ satisfies $\|A\| \leq 1 - \delta/2$.*

Proof. We verify that the Ising model has bounded spectral norm of A: note that each entry of A can be upper bounded by $|\beta|/2$ [Hay06], so a row sum of A is at most

$$\frac{|\beta|\Delta}{2} < \frac{(1-\delta)\Delta}{2} \ln\left(1 + \frac{2}{\Delta-2}\right) \leq \frac{(1-\delta)\Delta}{2}\left(\frac{2}{\Delta-2}\right) = (1-\delta)\left(1 + \frac{2}{\Delta-2}\right) < 1 - \delta/2,$$

where the last inequality holds for Δ large enough. □

We show next that Corollary 1.3 indeed follows from Theorem 3.1. For this, we first restate the corollary in a more precise manner.

Corollary 3.16. *Let $\delta \in (0, 1)$ and $\Delta \geq 3$. For the ferromagnetic Ising model with $\beta \leq (1-\delta)\beta_u(\Delta)$ on any graph G of maximum degree Δ and chromatic number χ, or for the ferromagnetic q-state Potts model with $q \geq 3$ and $\beta \leq \frac{2(1-\delta)}{\Delta}$ on the same graph, the mixing time of the SW dynamics satisfies*

$$T_{mix}(P_{SW}) = O\big(\chi \cdot \Delta^{\kappa} \cdot (\log n)^{1+\kappa}\big),$$

where $\kappa = 2 + \lceil \frac{4qe^2}{\delta} \rceil$.

Proof. If $\Delta = O(1)$, then the corollary was proved in a stronger form in [BCC+22]. Thus, we assume $\Delta \to \infty$.

We first show spectral independence. Let $q = 2$. Under the δ-uniqueness condition $0 < \beta < (1-\delta)\beta_u(\Delta)$, by Proposition 3.15 and Proposition 3.14, the Ising model μ_{Ising} satisfies $(4/\delta)$-spectral independence. For the q-state Potts model with $q \geq 3$, the Dobrushin influence matrix corresponding to μ_{Potts}

satisfies $\|A\| \le \frac{1}{2}\beta\Delta$; see proof of Theorem 2.13 in [Ull14]. Thus, if $\beta \le \frac{2(1-\delta)}{\Delta}$, then $\|A\| \le 1-\delta$, and by Proposition 3.14, μ_{Potts} satisfies $(2/\delta)$-spectral independence.

Letting $N(v)$ denote the neighborhood of v, and noting that for any configuration η on $N(v)$ we have $\mu(\sigma_v = c \mid \sigma_{N(v)} = \eta) \ge 1/(qe^2)$, we deduce that μ_{Potts} and μ_{Ising} are both $(1/(qe^2))$-marginally bounded. Therefore, by noting that $\kappa = 2 + \lceil \frac{4qe^2}{\delta} \rceil$ is a constant that only depends on δ, the mixing time bound follows from Theorem 3.1

$$T_{mix}(P_{SW}) = O\Big(\chi \cdot (C\Delta \log n)^{\kappa} \cdot (qe^2)^{(2+6\kappa)} (1+4/\delta)^{5\kappa} \cdot \log n\Big),$$

as desired. ◻

3.5.1 The SW dynamics on random graphs

As another application of Theorem 3.1, we consider the SW dynamics on a random graph generated from the classical $G(n, \frac{d}{n})$ model in which each edge is included independently with probability $p = d/n$; we consider the case where d is a constant independent of n. In this setting, while a typical graph has $\tilde{O}(n)$ edges, its maximum degree is of order $\Theta(\frac{\log n}{\log \log n})$ with high probability. Our results imply that the SW dynamics has polylogarithmic mixing on this type of graph provided β is small enough.

Corollary 3.17. *Let $\delta \in (0,1)$ and $d \in \mathbb{R}_{\ge 0}$ be constants independent of n. Suppose that $G \sim G(n, d/n)$ and G has maximum degree Δ. For the ferromagnetic Ising model with parameter $\beta < (1-\delta)\beta_u(\Delta)$ on G or the ferromagnetic q-Potts model with $q \ge 2$ and $\beta \le \frac{2(1-\delta)}{\Delta}$ on the same graph, the SW dynamics has $O\Big((\log n)^{5+2\lceil \frac{4qe^2}{\delta} \rceil}\Big)$ mixing time, with high probability over the choice of the random graph G.*

Corollary 3.17 is established using Corollary 3.16 and the following fact about random graphs.

Proposition 3.18 ([AN05])**.** *Let $G \sim G(n, \frac{d}{n})$ for a fixed $d \in \mathbb{R}_{\ge 0}$, and let χ be the chromatic number of G. With high probability over the choice of G, $\chi = k_d$ or $\chi = k_d + 1$, where k_d is the smallest integer k such that $d < 2k \log k$.*

Proof of Corollary 3.17. By Proposition 3.18, with high probability $G \sim G(n, \frac{d}{n})$ has chromatic number $\chi = O(d)$. Also, it is known that with high probability $\Delta = \Theta(\frac{\log n}{\log \log n})$. Suppose both properties hold. The result follows from Corollary 3.16. ◻

4 Systematic scan dynamics

In this section, we study the systematic scan dynamics for general spin systems (see Definition 2.1), which we define next. Given an ordering $\phi = [v_1, \dots, v_n]$ of the vertices, a systematic scan dynamics performs heat-bath updates on $v_1, \dots, v_n$ sequentially in this order. Recall that a heat-bath update on v_i simply means the replacement of the spin on v_i by a new spin assignment generated according to the conditional distribution in v_i given the configuration in $V \setminus \{v_i\}$. Let $P_i \in \mathbb{R}^{|\Omega| \times |\Omega|}$ be the transition matrix corresponding to a heat-bath update on the vertex v_i. The transition matrix of the systematic scan dynamics for the ordering ϕ can be written as $\mathcal{S}_\phi := P_n \dots P_1$. In general, $\mathcal{S}_\phi$ is not reversible, so as in earlier works we work with the symmetrized version of the scan dynamics that updates the spins in the order ϕ and in addition updates the spins in the reverse order of ϕ [Fil91, MT06]. The transition matrix of the symmetrized systematic scan dynamics can then be written as

$$P_\phi := \prod_{i=1}^{n} P_i \prod_{i=0}^{n-1} P_{n-i}.$$

Henceforth, we only consider the symmetrized version of the dynamics. Since P_ϕ is a symmetrized product of reversible transition matrices, one can straightforwardly verify its reversibility with respect to μ; its ergodicity follows from the assumption that the spin system is totally-connected (see Definition 2.2).

We show tight mixing time bounds for P_ϕ for monotone spin systems (see Definition 1.4). Our main result for the systematic scan dynamics is Theorem 1.6 from the introduction, which we restate here for convenience. The proof of this theorem is provided in Section 4.1.

Theorem 1.6. *Let $b > 0$, $\eta > 0$, and $\Delta \geq 3$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ. Let μ be the distribution of a totally-connected monotone spin system on G. If μ is η-spectrally independent and b-marginally bounded, then there exists a universal constant $C > 0$ such that for any ordering ϕ*

$$T_{mix}(P_\phi) = \Delta^{9+4\lceil \frac{2\eta}{b} \rceil} \cdot \left(\frac{C(\eta+1)^5}{b^6}\right)^{2+\lceil \frac{2\eta}{b} \rceil} \cdot O(\log n).$$

We complement Theorem 1.6 with a lower bound for the mixing time of systematic scan dynamics for a particular ordering ϕ. Specifically, on a bipartite graph $G = (V_E \cup V_O, E)$, an *even-odd scan dynamics* P_{EOE} is a systematic scan dynamics with respect to an ordering ϕ such that v_e appears before v_o in ϕ for all $v_e \in V_E$ and $v_o \in V_O$. In other words,

$$P_\phi = \prod_{i:v_i \in V_E} P_i \prod_{i:v_i \in V_O} P_i \prod_{i:v_i \in V_O} P_i \prod_{i:v_i \in V_E} P_i.$$

The above expression is well-defined without specifying the ordering in which the vertices in V_E and V_O are updated since the updates commute.

Lemma 4.1. *Let Δ be a constant and let G be an n-vertex connected bipartite graph with maximum degree Δ. The even-odd scan dynamics P_{EOE} for the ferromagnetic Ising model on G has mixing time $T_{mix}(P_{EOE}) = \Omega(\log n)$.*

The lower bound in Lemma 4.1 is proved in Section 4.2 using the machinery from [HS07] and the fact that even-odd scan dynamics does not propagate disagreements quickly (under a standard coupling). Our proof can thus be extended to other scan orderings that propagate disagreements slowly; however, there are orderings that do propagate disagreements quickly (think of a box in $\mathbb{Z}^2$ with the vertices sorted in a "spiral" from the boundary of the box to its center). For this type of ordering, the technique does not provide the $\Omega(\log n)$ lower bound. In addition, while we focus on the ferromagnetic Ising model to ensure clarity in the proof, the established lower bound is expected to apply to a broader class of spin systems.

4.1 Proof of main result for systematic scan dynamics: Theorem 1.6

The main technique in the proof of Theorem 1.6 is to compare the systematic scan dynamics with a fast mixing block dynamics via a censoring inequality developed in [FK13]. For this, we first introduce some notations and definitions.

We start by reviewing standard facts about the coupling method that will be used in our proofs; see [LPW06] for a more detailed background. A *coupling* of a Markov chain M specifies, for every pair of states $(X_t, Y_t) \in \Omega \times \Omega$ at every step t, a probability distribution P over (X_{t+1}, Y_{t+1}) such that when viewed in isolation, $\{X_t\}$ and $\{Y_t\}$ are valid instances of the chain M. The *optimal coupling lemma* says that for any two distributions μ and ν, we have

$$\|\mu - \nu\|_{TV} = \inf_{X\sim\mu, Y\sim\nu} \mathrm{P}\left[X \neq Y : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\right], \tag{30}$$

where the infimum is taken over all couplings of μ and ν. We focus on couplings of Markov chains such that if $X_s = Y_s$ then $X_t = Y_t$ for all $t \geq s$. Given a coupling P of M, the *coupling time*, is defined as

$$T_{coup}(M) := \min_{T>0}\left\{\max_{X_0\in\Omega, Y_0\in\Omega} \mathrm{P}[X_T \neq Y_T] \leq \frac{1}{4}\right\}.$$

It is a standard fact that for any coupling (X_t, Y_t), the coupling time bounds the mixing time as follows:

$$d(t) \leq \max_{X_0\in\Omega, Y_0\in\Omega} \mathrm{P}[X_T \neq Y_T], \text{ and thus } T_{mix}(M) \leq T_{coup}(M). \tag{31}$$

A coupling of two instances $\{X_t\}, \{Y_t\}$ of a Markov chain M is a *monotone coupling* if $X_{t+1} \geq_q Y_{t+1}$ whenever $X_t \geq_q Y_t$, where $\geq_q$ is the partial ordering of Ω. Let $\{X_{t,\sigma}\}$ denote the instance of M starting at configuration $\sigma \in \Omega$. If there exists a simultaneous monotone coupling of $\{X_{t,\sigma}\}$ for all $\sigma \in \Omega$ (i.e., a grand coupling), then we say M is a *monotone Markov chain*. It can be checked that P_ϕ is a monotone Markov chain for any ϕ (see e.g. [BCV20]).

We may also define a partial ordering $\preceq_\pi$ on the space of transition matrices. A function $f \in \mathbb{R}^{|\Omega|}$ is said to be *non-decreasing* if $f(\sigma) \geq f(\tau)$ whenever $\sigma \geq_q \tau$, or *non-increasing* if $f(\sigma) \leq f(\tau)$ whenever $\sigma \geq_q \tau$. We endow $\mathbb{R}^{|\Omega|}$ with the inner product $\langle f, g\rangle_\pi := \sum_{x\in\Omega} f(x)g(x)\pi(x)$, which induces a Hilbert space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle_\pi)$ denoted as $L_2(\pi)$. For transition matrices K and L whose stationary distributions are both π, we say $K \preceq_\pi L$ if $\langle Kf, g\rangle_\pi \leq \langle Lf, g\rangle_\pi$ for every non-negative and non-decreasing functions $f, g \in L_2(\pi)$. To show $K \preceq_\pi L$ in our applications, we use the following facts.

Proposition 4.2 ([FK13])**.** *Suppose π is the Gibbs distribution of a monotone spin system.*

1. *If $A_1 \preceq_\pi B_1$ and $A_2 \preceq_\pi B_2$, then for $0 \leq \lambda \leq 1$, $(1-\lambda)A_1 + \lambda A_2 \preceq_\pi (1-\lambda)A_1 + \lambda A_2$.*
2. *If $A_s \preceq_\pi B_s$ for $s = 1, \ldots, l$, then $A_1 \ldots A_l \preceq_\pi B_1 \ldots B_l$.*
3. *For any fixed v, let K_v be the heat-bath update at site v. Then, $K_v \preceq_\pi I$.*

Establishing such partial order between two transition matrices is significant as it would imply stochastic domination of the corresponding two chains (recall that for two distributions π and ν on Ω, we say π *stochastically dominates* ν, and denote as $\pi \succeq \nu$, if for any non-decreasing function $f \in \mathbb{R}^{|\Omega|}$, we have $\mathrm{E}_\pi[f] \geq \mathrm{E}_\nu[f]$). The following lemma captures such implication.

Lemma 4.3 ([FK13, BCV20])**.** *Suppose $\{X_t\}$ and $\{Y_t\}$ are monotone ergodic Markov chains reversible with respect to π, the Gibbs distribution of a monotone spin system. Let K_X and K_Y be the corresponding transition matrices of $\{X_t\}$ and $\{Y_t\}$. Suppose $K_X \preceq_\pi K_Y$. Then $X_t \preceq Y_t$ for all $t \geq 0$ if the initial states X_0 and Y_0 are sampled from a common distribution ν such that ν/π is non-decreasing; if ν/π is instead non-increasing, then $Y_t \preceq X_t$ for all $t \geq 0$, where $\preceq$ as a relation for X_t and Y_t denotes stochastic domination of their corresponding distributions at time t.*

We now provide our proof of Theorem 1.6.

Proof of Theorem 1.6. We partition V into k disjoint independent sets $I_1, I_2, \ldots, I_k$, where $k = O(\Delta)$. Set $\mathcal{B} = \{I_1, \ldots, I_k\}$ and define $P_{\mathcal{B}}$ to be the heat-bath block dynamics w.r.t. these independent sets. Fix an ordering $\phi = [v_1, \ldots, v_n]$, and fix $j \in \{1, \ldots, k\}$. Let K_j be the transition matrix corresponding to heat-bath update in the independent set I_j, which can also be seen as a systematic scan on I_j according to the ordering defined by ϕ. We define $\hat{P}_i$ to be P_i if $i \in I_j$ and the identity matrix I otherwise so that

$$K_j = K_j^2 = \left(\prod_{i:v_i\in I_j} P_i\right)^2 = \left(\prod_{i:v_i\in I_j} P_i \prod_{i:v_i\notin I_j} I\right)^2 = \prod_{i=1}^{n} \hat{P}_i \prod_{i=0}^{n-1} \hat{P}_{n-i}.$$

Note that in the computation above, P_i and $P_{i'}$ commute for $v_i, v_{i'} \in I_j$, and I commutes with arbitrary matrices. By Proposition 4.2(3), we obtain $P_i \preceq_\mu \hat{P}_i$ for all i, and hence by Proposition 4.2(2), we obtain $P_\phi \preceq_\mu K_j$ for any j, and consequently, by Proposition 4.2(1),

$$P_\phi \preceq_\mu \frac{1}{k}\sum_{j=1}^{k} K_j = P_{\mathcal{B}}. \tag{32}$$

Let $+$ and $-$ denote the top and the bottom elements in $[q]$ respectively. Let $\{X_t^+\}$ (resp., $\{X_t^-\}$) be an instance of a Markov chain with transition matrix P_ϕ starting from the all $+$ (resp., all $-$) configuration. Similarly, let $\{Y_t^+\}$ (resp., $\{Y_t^-\}$) be an instance of $P_{\mathcal{B}}$ starting from the all $+$ (resp., all $-$) configuration. P_ϕ is monotone, so we can define a grand monotone coupling of $\{X_t^+\}$ and $\{X_t^-\}$ such that $X_t^- \preceq_q X_t^+$ for all $t \geq 0$, which with (31) further implies that the mixing time of a systematic scan can be upper bounded by the coupling time of the all $+$ and all $-$ configurations.

Letting ν^+ and (resp., ν^-) denote the trivial distribution concentrated on the all $+$ (resp., all $-$) configuration, we note that ν^+/μ is non-decreasing and ν^-/μ is non-increasing. Then Lemma 4.3 and (32) imply that for all $t \geq 0$,

$$Y_t^- \preceq X_t^- \preceq X_t^+ \preceq Y_t^+.$$

For any $v \in V$ and all $t \geq 0$, $X_t^- \preceq_q X_t^+$ implies that

$$\begin{aligned}
\Pr[X_t^+(v) \neq X_t^-(v)] &\leq \sum_{c\in[q]} \Pr[X_t^+(v) \geq c, X_t^-(v) < c] \\
&= \sum_{c\in[q]} \Pr[X_t^+(v) \geq c] - \Pr[X_t^-(v) \geq c].
\end{aligned}$$

Then, since $Y_t^- \preceq X_t^-$ and $X_t^+ \preceq Y_t^+$, we obtain that

$$\begin{aligned}
\sum_{c\in[q]} \Pr[X_t^+(v) \geq c] - \Pr[X_t^-(v) \geq c] &\leq \sum_{c\in[q]} \Pr[Y_t^+(v) \geq c] - \Pr[Y_t^-(v) \geq c] \\
&\leq \sum_{c\in[q]} \left|\Pr[Y_t^+(v) \geq c] - \Pr[Y_t^-(v) \geq c]\right| \\
&\leq q\|P_{\mathcal{B}}^t(+,\cdot) - P_{\mathcal{B}}^t(-,\cdot)\|_{TV} \\
&\leq q\left(\|P_{\mathcal{B}}^t(+,\cdot) - \mu(\cdot)\|_{TV} + \|P_{\mathcal{B}}^t(-,\cdot) - \mu(\cdot)\|_{TV}\right). \qquad (33)
\end{aligned}$$

Since μ is η-spectrally independent and b-marginally bounded, it follows from Theorem 3.3 and Remark 2 that $P_{\mathcal{B}}$ satisfies the relative entropy decay with rate

$$r \geq \frac{b^{6\kappa}}{k\Delta^{4\kappa}\cdot(C(\eta+1)^5)^\kappa}, \tag{34}$$

where $\kappa = 2 + \lceil\frac{2\eta}{b}\rceil$. Let $b' := (C(\eta+1)^5/b^6)^\kappa$, and let

$$T := k\Delta^{4\kappa}b'\log\left(\frac{\log(\mu_{min}^{-1})}{1/(4qn)}\right) = O(\Delta^{4\kappa+1}b'\log(qn)).$$

By (5) and (34), $T_{mix}(P_{\mathcal{B}}, 1/(8qn)) \leq T$. Then for any $\sigma \in \Omega$,

$$\|P_{\mathcal{B}}^T(\sigma,\cdot) - \mu(\cdot)\|_{TV} \leq \frac{1}{8qn},$$

so we have $\Pr[X_T^+(v) \neq X_T^-(v)] \leq 1/(4n)$. By a union bound, $\Pr[X_T^+ \neq X_T^-] \leq 1/4$, and therefore

$$T_{mix}(P_\phi) \leq T = O(\Delta^{4\kappa+1}b'\log(qn)),$$

establishing the desired bound for the mixing time. □

4.2 Proof of the lower bound: Lemma 4.1

We provide next the proof of Lemma 4.1. Our proof extends the argument from [HS07] for the Glauber dynamics and also uses ideas from [BCSV23, BCP+22]. The following fact will be used in our proof.

Lemma 4.4 (Lemma 35, [BCP+22]). *Let $\{X_t\}$ denote a discrete-time Markov chain with finite state space Ω, reversible with respect to π and with a positive semidefinite transition matrix. Let $B \subseteq \Omega$ denote an event. If X_0 is sampled proportional to π on B, then $\Pr[X_t \in B] \geq \pi(B)$ for all $t \geq 0$, and for all $t \geq 1$,*

$$\Pr[X_t \in B] \geq \pi(B) + (1 - \pi(B))^{-t+1} \left[\Pr(X_1 \in B) - \pi(B)\right]^t .$$

We can now prove Lemma 4.1.

Proof of Lemma 4.1. Suppose n is sufficiently large. Let $R = \lceil \frac{\ln n}{8 \ln \Delta} \rceil$ and let $T = \alpha \ln n < R/3$ for some $\alpha > 0$ we will specify later. We will show that for some (random) starting configuration $X_0 \in \Omega$,

$$\|\mu_{\text{Ising}}(\cdot) - P_{EOE}^T(X_0, \cdot)\|_{TV} > 1/4, \tag{35}$$

and hence by definition $T_{mix}(P_{EOE}) \geq T$. As G has maximum degree Δ, we can always find a subset $V_C \subseteq V$ of size at least $n^{1/4}$ whose pairwise graph distances are at most $2R$. Let $G_C := \cup_{u \in V_C} B(u, R)$. We consider a restriction of the even-odd scan dynamics on G_C. Let $\{X_t\}$ be an instance of the even-odd scan dynamics, and let $\{Y_t\}$ be an even-odd scan dynamics that only updates spins for vertices in G_C, starting from the same configuration as $\{X_t\}$ which will be specified next.

Let $N := n^{1/4}$, and let $f : \Omega \to \mathbb{R}$ be the function given by $f(\sigma) = \frac{1}{N} \sum_{v \in V_C} \mathbb{1}(\sigma(v) = +1)$. To show (35), it suffices to find a distribution for $X_0 \in \Omega$ and a threshold $A \in \mathbb{R}$ such that

$$\left|\Pr\left[f(X_T) \geq A\right] - \Pr_{\sigma \sim \mu_{\text{Ising}}}\left[f(\sigma) \geq A\right]\right| > 1/4. \tag{36}$$

We define X_0 by setting the configuration on $V_C \cup (V \setminus G_C)$ to be the all $+1$ configuration and for each $v_C \in V_C$ sampling the configuration in $B(v_C, R) \setminus \{v_C\}$ conditional on the all $+1$ configuration on $V_C \cup (V \setminus G_C)$. Let π denote the conditional distribution on G_C with a fixed all $+1$ configuration on $V \setminus G_C$. Define $A := \mathrm{E}_{\sigma \sim \pi}\left[f(\sigma)\right] + N^{-1/3}$. We will show next that

1. $\Pr_{\sigma \sim \mu_{\text{Ising}}}\left[f(\sigma) \geq A\right] \leq 1/2$;
2. Under the identity coupling, $f(X_t) = f(Y_t)$ for $t \leq T$. The identity coupling is the standard coupling that updates the same vertex in both chains at the same time and maximizes the probability that the spin value at the vertex agrees after the update;
3. $\Pr\left[f(Y_T) \geq A\right] > \frac{3}{4}$,

and thus (36) follows.

We first give the upper bound for $\Pr_{\sigma \sim \mu_{\text{Ising}}}\left[f(\sigma) \geq A\right]$. Since the ferromagnetic Ising model is monotone, and f is a non-decreasing function, for any boundary condition τ on $\Omega_{V \setminus G_C}$,

$$\mathrm{E}_{\sigma \sim \pi}\left[f(\sigma)\right] \geq \mathrm{E}_{\sigma \sim \mu_{\text{Ising}}^\tau}\left[f(\sigma)\right].$$

For any $\tau \in \Omega_{V \setminus G_C}$, if σ is generated from μ_{Ising}^τ, then $f(\sigma)$ is the average of N independent indicator random variables. By Hoeffding's inequality,

$$\Pr_{\sigma \sim \mu_{\text{Ising}}^\tau}\left[f(\sigma) \geq A\right] \leq \Pr_{\sigma \sim \mu_{\text{Ising}}^\tau}\left[f(\sigma) \geq \mathrm{E}_{\sigma \sim \mu_{\text{Ising}}^\tau}\left[f(\sigma)\right] + N^{-1/3}\right] \leq \exp\left(-\frac{2 \cdot N^{4/3}}{N}\right) < \frac{1}{2},$$

and thus

$$\Pr_{\sigma\sim\mu_{\text{Ising}}}\left[f(\sigma)\geq A\right]=\sum_{\tau\in\Omega_{V\setminus G_C}}\Pr_{\sigma\sim\mu_{\text{Ising}}^{\tau}}\left[f(\sigma)\geq A\right]\cdot\mu(\tau)<\frac{1}{2}.$$

To see that $f(X_t)=f(Y_t)$, we consider the speed of "disagreement propagation". Note that $f(X_0)=f(Y_0)$ since $X_0=Y_0$. The key observation is that under the identity coupling, in one step of the coupled even-odd scan dynamics, the disagreement at any vertex v can be propagated only to vertices at distance at most 3 from v. Since $R>3T$, we can guarantee that $X_t(v)=Y_t(v)$ for all $v\in V_C$ and all $t\leq T$.

Finally, we provide a bound for $\Pr\left[f(Y_T)\geq A\right]$. Fix $v\in V_C$. Let π_v denote the Ising model distribution restricted to $B(v,R)$ under the all +1 boundary condition outside of $B(v,R)$. Note that $\bigotimes_{v\in V_C}\pi_v=\pi$. Let $\{Y_t^v\}$ denote the Markov chain obtained by projecting $\{Y_t\}$ to $B(v,R)$. Since the boundary of $B(v,R)$ is fixed, $\{Y_t^v\}$ is simply an even-odd scan dynamics on $B(v,R)$ under the all +1 boundary condition. It can be checked that $\{Y_t^v\}$ is reversible with respect to π_v and that it has a positive semidefinite transition matrix. We define $\mathcal{B}_v$ to be the event (or subset of configurations) that v is assigned spin +1. It can also be verified that μ_{Ising} is b-marginally bounded for some constant $b=b(\beta,\Delta)$, so $b\leq\pi_v(\mathcal{B}_v)\leq 1-b$. Moreover, we have the following fact, which we prove later.

Claim 4.5. *There exists a constant $c:=c(\beta,\Delta)>0$ such that* $\Pr(Y_1^v\in\mathcal{B}_v)>\pi_v(\mathcal{B}_v)+c$.

By Lemma 4.4 and Claim 4.5, for all $t\geq 1$,

$$\Pr[Y_t^v\in\mathcal{B}_v]\geq\pi_v(\mathcal{B}_v)+b^{-t+1}\left[\Pr(Y_1^v\in\mathcal{B}_v)-\pi_v(\mathcal{B}_v)\right]^t\geq\pi_v(\mathcal{B}_v)+\frac{c^t}{b^{t-1}}.$$

Using this and the definition of f, we have

$$\mathrm{E}[f(Y_T)]=\frac{1}{N}\sum_{u\in V_C}\Pr[Y_T^u\in\mathcal{B}_u]\geq\frac{1}{N}\sum_{u\in V_C}\left(\pi_u(\mathcal{B}_u)+\frac{c^T}{b^{T-1}}\right)=\mathrm{E}_{\sigma\sim\pi}\left[f(\sigma)\right]+\frac{c^T}{b^{T-1}}.$$

Set $T:=\min(\frac{R}{3},\frac{\frac{1}{12}\ln n-\ln\frac{2}{b}}{\ln\frac{b}{c}})$, so that $\frac{c^T}{b^{T-1}}\geq 2N^{-1/3}$. Thus, $\mathrm{E}[f(Y_T)]\geq A+N^{-1/3}$. By Hoeffding's inequality, we obtain

$$\Pr\left[f(Y_T)<A\right]\leq\Pr\left[f(Y_T)<\mathrm{E}[f(Y_T)]-N^{-1/3}\right]\leq\exp\left[-\frac{2N^{4/3}}{N}\right]<\frac{1}{4}.$$

Therefore, the mixing time of P_{EOE} is at least $T=\Omega(\log n)$. $\square$

It remains to prove Claim 4.5.

Proof of Claim 4.5. Let P be the even-odd dynamics defined on $V'=B(v,R)$, and suppose $V'=V_E\cup V_O$ is a connected bipartite graph. Suppose $v\in V_O$ without loss of generality. Recall that the transition matrix of P is

$$\prod_{i:v_i\in V_E}P_i\prod_{i:v_i\in V_O}P_i\prod_{i:v_i\in V_E}P_i.$$

We use Y_E, Y_{OE} and $Y_{EOE}=Y_1^v$ to denote the configuration of Y_0^v after the updates $\prod_{i:v_i\in V_E}P_i$ on even vertices for the first time, after the updates $\prod_{i:v_i\in V_O}P_i$ on odd vertices and after update $\prod_{i:v_i\in V_E}P_i$ respectively. Since the last set of updates on the even vertices do not affect the spin at v, we have

$$\Pr(Y_1^v\in\mathcal{B}_v)=\mathrm{E}\big[\mathbb{1}(Y_{EOE}\in\mathcal{B}_v)\big]=\mathrm{E}\big[\mathbb{1}(Y_{OE}\in\mathcal{B}_v)\big]=\mathrm{E}\left[\mathrm{E}\big[\mathbb{1}(Y_{OE}\in\mathcal{B}_v)\mid Y_E\big]\right].$$

Let $N(w)$ denote the set of vertices in V' adjacent to w. For a configuration $\sigma\in\Omega$ and $w\in V$, we define $S(\sigma;w):=\sum_{x\in N(w)}\mathbb{1}(\sigma_x=+1)$ and $g_w:\mathbb{Z}\to[0,1]$ given by $g_w(y):=\mu_{\text{Ising}}(\sigma_w=+1\mid S(\sigma;w)=y)$. Let

π_v^+ (resp. π_v^-) be distribution on V' given by $\pi_v^+(\sigma) = \pi_v(\sigma \mid \sigma \in \mathcal{B}_v)$ (resp. $\pi_v^-(\sigma) = \pi_v(\sigma \mid \sigma \notin \mathcal{B}_v)$). Recall that Y_0^v is a configuration drawn from π_v^+ and by noting that

$$\pi_v^+ \cdot \left(\prod_{i: v_i \in V_E} P_i \right) = \pi_v^+,$$

so Y_E can also be viewed as a configuration drawn from π_v^+. Hence, by the definition of the Gibbs update, we have

$$\mathrm{E}\left[\mathrm{E}\left[\mathbb{1}(Y_{OE} \in \mathcal{B}_v) \mid Y_E\right]\right] = \mathrm{E}_{\tau \sim \pi_v^+}\left[g_v(S(\tau, v))\right].$$

Similarly,

$$\pi_v(\mathcal{B}_v) = \mathrm{E}_{\sigma \sim \pi_v}\left[g_v(S(\sigma, v))\right].$$

By Strassen's theorem, there exists a coupling of (σ, τ) such that $\sigma \sim \pi_v, \tau \sim \pi_v^+$ and $\sigma \leq_q \tau$. Then $\sigma_{N(v)} \neq \tau_{N(v)}$ implies $S(\tau, v) \geq S(\sigma, v) + 1$. Therefore,

$$\begin{aligned}
\Pr(Y_1^v \in \mathcal{B}_v) - \pi_v(\mathcal{B}_v) &= \mathrm{E}_{\tau \sim \pi_v^+}\left[g_v(S(\tau, v))\right] - \mathrm{E}_{\sigma \sim \pi_v}\left[g_v(S(\sigma, v))\right] \\
&= \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[g_v(S(\tau, v)) - g_v(S(\sigma, v))\right] \\
&\geq \min_{i \leq \deg(v)} (g_v(i, v) - g_v(i-1, v)) \cdot \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[S(\tau, v)) - S(\sigma, v)\right] \\
&\geq \min_{i \leq \deg(v)} (g_v(i, v) - g_v(i-1, v)) \cdot \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[\mathbb{1}(\sigma_{N(v)} \neq \tau_{N(v)})\right].
\end{aligned}$$

It can be checked that $\min_{i \leq \deg(v)}(g_v(i, v) - g_v(i-1, v)) \geq c_2$, where $c_2 := c_2(\beta, \Delta)$, Moreover, for any $u \in N(v)$ we have

$$\mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[\mathbb{1}(\sigma_{N(v)} \neq \tau_{N(v)})\right] \geq \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[\mathbb{1}(\sigma_u \neq \tau_u)\right].$$

Fix u and let $\Lambda := V' \setminus \{u, v\}$. Since $\sigma_u \leq \tau_u$, $\sigma_u \neq \tau_u$ implies that $\sigma_u = -1$ and $\tau_u = +1$. Thus we obtain

$$\begin{aligned}
\mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[\mathbb{1}(\sigma_u \neq \tau_u)\right] &= \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[\mu_{\mathrm{Ising}}(\tau_u = +1 \mid \tau_\Lambda) - \mu_{\mathrm{Ising}}(\sigma_u = +1 \mid \sigma_\Lambda)\right] \\
&= \mathrm{E}_{(\sigma,\tau) \sim (\pi_v, \pi_v^+)}\left[g_u(S(\tau, u)) - g_u(S(\sigma, u))\right] \\
&\geq b \cdot \mathrm{E}_{(\sigma,\tau) \sim (\pi_v^-, \pi_v^+)}\left[g_u(S(\tau, u)) - g_u(S(\sigma, u))\right],
\end{aligned}$$

where the inequality is due to the b-bounded marginal condition of μ_{Ising} which requires $\sigma_v = -1$ with probability at least b. Note that if $\sigma \sim \pi_v^-, \tau \sim \pi_v^+$ and $\sigma \leq_q \tau$, then $S(\tau, u) \geq S(\sigma, u) + 1$. Hence,

$$\mathrm{E}_{(\sigma,\tau) \sim (\pi_v^-, \pi_v^+)}\left[g_u(S(\tau, u)) - g_u(S(\sigma, u))\right] \geq \min_{i \leq \deg(u)} (g_u(i, u) - g_u(i-1, u)) > c_3,$$

for some $c_3 = c_3(\beta, \Delta) > 0$. Therefore, we established that

$$\Pr(Y_1^v \in \mathcal{B}_v) - \pi_v(\mathcal{B}_v) \geq c_2 c_3 b,$$

and $c_2 c_3 b$ depends only on β, Δ. □

4.3 Applications of Theorem 1.6

We discuss next some applications of Theorem 1.6. As a first application, we can establish *optimal* mixing for the systematic scan dynamics on the ferromagnetic Ising model under the δ-uniqueness condition, improving the best known results that hold under the Dobrushin-type conditions [SIM93, DGJ06a, Hay06]. This result was stated in Corollary 1.7 in the introduction and is proved next. For this, we recall that under δ-uniqueness condition, the Ising distribution μ_{Ising} satisfies spectral independence and the bounded marginals condition.

Proposition 4.6 ([CLV20, CLV21]). *The ferromagnetic Ising model with parameter β such that $\bar{\beta}_u(\Delta)(1-\delta) < \beta < \beta_u(\Delta)(1-\delta)$ is $O(1/\delta)$-spectrally independent and b-marginally bounded with $b = O(1)$.*

Proof of Corollary 1.7. We fix $\delta \in (0,1)$ and first assume that Δ is a constant. By Proposition 4.6, the ferromagnetic Ising model with parameter $\beta < (1-\delta)\beta_u(\Delta)$ satisfies η-spectral independence and b-bounded marginals, where $\eta = O(1/\delta)$ and b is a constant. Since the ferromagnetic Ising model is a monotone system, it follows from Theorem 1.6 that $T_{mix} = O(\log n)$ for any ordering ϕ.

Now, when $\Delta \to \infty$ as $n \to \infty$, by Proposition 3.15, the Dobrushin's influence matrix A of ferromagnetic Ising model satisfies that $\|A\| \le 1 - \delta/2$. Under this assumption, it is known that $T_{mix} = O(\log n)$ for any ordering ϕ; see [Hay06]. □

We can similarly show mixing time bound for the systematic scan dynamics of the hardcore model on bipartite graphs under δ-uniqueness condition.

Corollary 4.7. *Let $\delta \in (0,1)$ be a constant. Suppose G is an n-vertex bipartite graph of maximum degree $\Delta \ge 3$. For the hardcore model on G with fugacity λ such that $0 < \lambda < (1-\delta)\lambda_u(\Delta)$, where $\lambda_u(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$ is the tree uniqueness threshold on the Δ-regular tree, the systematic scan with respect to any ordering ϕ satisfies*

$$T_{mix}(P_\phi) = \Delta^{O(1/\delta)} \cdot O(\log n).$$

Proof of Corollary 4.7. The hardcore model on a bipartite graph $(V_1 \cup V_2, E)$ with fugacity $0 < \lambda < (1-\delta)\lambda_u(\Delta)$ is monotone, and [CLV21, AJK+22, CLY23] show that it satisfies $O(1/\delta)$-spectral independence and the $\Omega(\lambda)$-bounded marginals condition. Theorem 1.6 then implies $\Delta^{O(1/\delta)} \cdot O(\log n)$ mixing of systematic scan for any ordering. □

We consider next the application of Theorem 1.6 to the special case where the underlying graph is a cube of the d-dimensional lattice graph $\mathbb{Z}^d$. We show that strong spatial mixing implies optimal $O(\log n)$ mixing of any systematic scan dynamics. Previously, under the same type of condition, [BCSV19] gave an $O(\log n (\log\log n)^2)$ mixing time bound for arbitrary orderings, and an $O(\log n)$ mixing time bound for a special class of scans that (deterministically) propagate disagreements slowly under the standard identity coupling. We first provide the definition of our SSM condition.

Definition 4.8. We say a spin system μ on $\mathbb{Z}^d$ satisfies the *strong spatial mixing (SSM)* condition if there exist constants $\alpha, \gamma, L > 0$ such that for every d-dimensional rectangle $\Lambda \subset \mathbb{Z}^d$ of side length between L and $2L$ and every subset $B \subset \Lambda$, with any pair (τ, τ') of boundary configurations on $\partial\Lambda$ that only differ at a vertex u, we have

$$\|\mu_B^\tau(\cdot) - \mu_B^{\tau'}(\cdot)\|_{TV} \le \gamma \cdot \exp(-\alpha \cdot dist(u, B)),$$

where $dist(\cdot,\cdot)$ denotes graph distance.

The definition above differs from other variants of SSM in the literature (e.g., [DSVW04, BCSV19, MOS94]) in that Λ has been restricted to "regular enough" rectangles. In particular, our variant of SSM is easier to satisfy than those in [DSVW04, MOS94] but more restricting than the one in [BCSV19] (that only considers squares). Nevertheless, it follows from [CP21, MOS94, Ale98, BDC12] that for the ferromagnetic Ising model, this form of SSM holds up to a critical threshold temperature $\beta < \beta_c(2) = \ln(1+\sqrt{2})$ on $\mathbb{Z}^2$.

Corollary 1.9 from the introduction states that for b-marginally bounded monotone spin system on d-dimensional cubes $V \subseteq \mathbb{Z}^d$, SSM implies that the mixing time of any systematic scan P_ϕ is $O(\log n)$. As mentioned there, this result in turn implies that any systematic scan dynamics for the ferromagnetic Ising model is mixing in $O(\log n)$ steps on boxes of $\mathbb{Z}^2$ when $\beta < \beta_c(2)$. Another interesting consequence of Corollary 1.9 is that we obtain $O(\log n)$ mixing time for any systematic scan dynamics P_ϕ for the hardcore model on $\mathbb{Z}^2$ when $\lambda < 2.538$, which is the best known condition for ensuring SSM [SSSY17, RST+13].

Our proof of Corollary 1.9 relies on Lemma 1.8 that is restated below. Remarkably, Lemma 1.8 generalizes beyond monotone systems and may be of independent interests.

Lemma 1.8. *For a spin system on a d-dimensional cube $V \subseteq \mathbb{Z}^d$, SSM implies η-spectral independence, where $\eta = O(1)$.*

Proof of Corollary 1.9. Assume a monotone spin system satisfies SSM condition. Then the spin system satisfies η-spectral independence, where $\eta = O(1)$ by Lemma 1.8. By noting that $\Delta = 2^d$ the corollary follows from Theorem 1.6. $\square$

Lastly, we give a proof of Lemma 1.8. For this, we recall the notion of a κ-contractive coupling which is known to imply spectral independence. We say a distribution μ is *κ-contractive* with respect to a Markov chain P if for all $X_0, Y_0 \in \Omega$, there exists a coupling of step of P so that

$$\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0] \leq \kappa d(X_0, Y_0),$$

where $d(\cdot, \cdot)$ denotes the Hamming distance of two configurations. The following lemma from [BCC+22] shows that spectral independence follows from the existence of a contractive coupling with respect to a heat-bath block dynamics.

Lemma 4.9 ([BCC+22])**.** *If μ is κ-contractive with respect to a block dynamics, then μ is $(\frac{2DM}{1-\kappa})$-spectrally independent, where M is the maximum block size and D is the maximum probability of a vertex being selected as part of a block in any step of the block dynamics.*

With this lemma on hand, we can now prove Lemma 1.8.

Proof of Lemma 1.8. Let L be a sufficiently large constant so that the SSM condition is satisfied; we will choose L later. Let V be a d-dimensional cube of $\mathbb{Z}^d$. We define a heat-bath block dynamics $P_{\mathcal{B}}$ with respect to a collection $\mathcal{B}$ of d-dimensional rectangles in V. Precisely, let $S_v := \{w \in \mathbb{Z}^d : d_\infty(w, v) < L\}$, and let $\mathcal{B}$ be the set of blocks $\{S_v \cap V\}_{v \in V}$. Given a configuration X_t, the heat-bath block dynamics $P_{\mathcal{B}}$ obtains a configuration X_{t+1} in 3 steps as follows:

1. Choose $v \in V$ uniformly at random. Let $S'_v := S_v \cap V$.
2. Generate a configuration $\sigma \in \Omega_{S'_v}$ from $\mu^\tau_{S'_v}(\cdot)$, where $\tau \in \Omega_{V \setminus S'_v}$ is given by $\tau(u) = X_t(u)$;
3. Let $X_{t+1}(u) = \sigma(u)$ if $u \in S'_v$ and $X_{t+1}(u) = X_t(u)$ otherwise.

We will show that μ is κ-contractive with respect to $P_{\mathcal{B}}$ whenever SSM holds. Our argument builds upon [DSVW04] but works for $P_{\mathcal{B}}$ under our weaker form of SSM condition, in which the geometry is restricted to d-dimensional rectangles of large side lengths. One can verify that if $\Lambda = S_v \cap V \in \mathcal{B}$, then Λ is a d-dimensional rectangle of side lengths between L and $2L$. The argument in [DSVW04] requires a stronger form of SSM to deal with the set of blocks $\mathcal{B}' = \{\Lambda = S_v \cap V : \Lambda \neq \emptyset, v \in \mathbb{Z}^d\}$ which contains arbitrarily thin rectangles, and this stronger form of SSM condition does not hold up to β_c for the ferromagnetic Ising.

Fix (X_0, Y_0) such that there exists exactly one vertex $u \in V$ such that $X_0(u) \neq Y_0(u)$ and $X_0(v) = Y_0(v)$ for all $v \neq u$. We select the same $v \in V$ in the first step of $P_{\mathcal{B}}$ in both chains; let $\Lambda = S'_v$. There are three cases with regard to the position of the disagreeing vertex u: u is contained in Λ, u is on the boundary of Λ, or u is far from Λ. Let $\partial\Lambda$ denote the external boundary of Λ. If $u \in \Lambda$ or $u \notin (\Lambda \cup \partial\Lambda)$, since the boundary conditions are identical, we generate the same configuration $\sigma \sim \mu^\tau_\Lambda$ to update Λ in both chains such that $X_1(\Lambda) = Y_1(\Lambda)$, where $\tau := X_0(\partial\Lambda) = Y_0(\partial\Lambda)$. Hence, $\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \in \Lambda] = 0$ and $\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \notin (\Lambda \cup \partial\Lambda)] = 1$.

It remains to define the coupling in the case when $u \in \partial\Lambda$, and we would need an upper bound for $\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \in \partial\Lambda]$. For this, we use the SSM condition. Let $B := \{w \in \Lambda : d(w, u) \geq r\}$, where $r := \frac{1}{2}\left(\frac{L}{d}\right)^{1/2d}$, and let τ and τ' be the boundary conditions of Λ in X_0 and Y_0 respectively. By assumption, τ and τ' are only different at u. We can view the coupling of the update on Λ as consisting of three steps:

1. Generate two configurations $\sigma_1, \sigma_2 \in \Omega_B$ from μ_B^{τ} and $\mu_B^{\tau'}$ using the optimal coupling of the two distributions;
2. Independently generate two configurations $\sigma_3, \sigma_4 \in \Omega_{\Lambda\setminus B}$ from $\mu_{\Lambda\setminus B}^{\tau\cup\sigma_1}$ and $\mu_{\Lambda\setminus B}^{\tau'\cup\sigma_2}$;
3. Let $X_1(u) = \sigma_1(u)$ and $Y_1(u) = \sigma_2(u)$ if $u \in B$, and $X_1(u) = \sigma_3(u)$ and $Y_1(u) = \sigma_4(u)$ if $u \in \Lambda \setminus B$.

Clearly, $X_1(\Lambda) \sim \mu_\Lambda^\tau$ and $Y_1(\Lambda) \sim \mu_\Lambda^{\tau'}$, so the coupling is valid. By (30), there exists a coupling P used for the first step such that

$$\mathrm{P}[\sigma_1 \neq \sigma_2] = \|\mu_B^\tau - \mu_B^{\tau'}\|_{TV}.$$

Moreover, SSM implies that there exist constants $\gamma, \alpha > 0$ such that

$$\|\mu_B^\tau - \mu_B^{\tau'}\|_{TV} \leq \gamma \cdot \exp(-\alpha \cdot dist(u, B)) \leq \gamma \cdot e^{-\alpha r}.$$

Also, $|\Lambda| \leq (2L)^d$ and $|\Lambda \setminus B| \leq (2r)^d$. Put together, we have

$$\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \in \partial\Lambda] \leq 1 + |\Lambda \setminus B| + |\Lambda| \cdot \mathrm{P}[\sigma_1 \neq \sigma_2] \leq 1 + (2r)^d + (2L)^d \cdot \gamma \cdot e^{-\alpha r}.$$

Let $N := |\mathcal{B}|$. Therefore, by noting that $\Pr[u \notin \Lambda] \geq L^d$ we obtain

$$\begin{aligned}
\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0] &= \mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \in \partial\Lambda] \cdot \Pr[u \in \partial\Lambda] + \mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \in \Lambda] \cdot \Pr[u \in \Lambda] \\
&\quad + \mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0, u \notin (\Lambda \cup \partial\Lambda)] \cdot \Pr[u \notin (\Lambda \cup \partial\Lambda)] \\
&\leq 1 + \Pr[u \in \partial\Lambda] \cdot [(2r)^d + (2L)^d \cdot \gamma \cdot e^{-\alpha r}] - \Pr[u \in \Lambda] \\
&\leq 1 + \frac{2d \cdot (2L)^{d-1}}{N} \cdot [(2r)^d + (2L)^d \cdot \gamma \cdot e^{-\alpha r}] - \frac{L^d}{N} \\
&= 1 + \frac{L^{d-1}}{N} \cdot \left[2^d d \cdot \left(\sqrt{\frac{L}{d}} + \frac{(2L)^d \cdot \gamma}{\exp(\alpha \cdot \sqrt[2d]{\frac{L}{d}})}\right) - 2L\right].
\end{aligned} \tag{37}$$

Recall that $N = O(n)$. By choosing $L = L(d, \alpha, \gamma)$ sufficiently large, we obtain

$$\mathbb{E}[d(X_1, Y_1) \mid X_0, Y_0] \leq 1 - \frac{L^{d-1}}{N} = 1 - \Omega\left(\frac{1}{N}\right) = 1 - \Omega\left(\frac{1}{n}\right).$$

In the case where blocks are of maximum size $(2L)^d$ and where each vertex is covered by at most $(2L)^d$ number of blocks at any step, $D = \Theta(n^{-1})$ and $M = O(1)$. Thus, Lemma 4.9 implies that μ is η-spectrally independent, where

$$\eta = \frac{\Theta(n^{-1})}{1 - (1 - \Omega(n^{-1}))} = O(1),$$

as desired. □

5 General block dynamics

In this section, we give an upper bound for the mixing time of the block dynamics of a totally-connected spin system on general graphs. In particular, we prove Theorem 1.10 from the introduction.

We present next a more general form of entropy factorization. In particular, KPF and UBF are special cases of it. A Gibbs distribution μ is said to satisfy the *general block factorization of entropy (GBF)* with constant $C_{\rm GBF}$ if for all functions $f : \Omega \to \mathbb{R}_{\geq 0}$, and for all probability distributions α over the set of all subsets of V,

$$\alpha_{min} \cdot \mathrm{Ent}_\mu(f) \leq C_{\rm GBF} \sum_{U \subseteq V} \alpha(U) \mathrm{E}_{\tau \sim \mu_{V \setminus U}} \left[\mathrm{Ent}_U^\tau(f^\tau)\right],$$

where $\alpha_{min} = \min_{v \in V} \sum_{U : v \in U} \alpha(U)$. The notion of GBF is closely related to the general block dynamics [CP21, BCC⁺22, CMT14]. Indeed, the following proposition shows that a bound for $C_{\rm GBF}$ yields a bound for the modified log-Sobolev constant of general block dynamics.

Proposition 5.1 (Lemma 2.8 in [BCC⁺22])**.** *If the Gibbs distribution μ of a spin system is totally-connected and satisfies GBF with constant $C_{\rm GBF}$, then the general block dynamics $P_{\mathcal{B},\alpha}$ w.r.t. $(\mathcal{B}, \alpha)$ satisfies relative entropy decay with rate at least $\frac{\alpha_{min}}{C_{\rm GBF}}$ and satisfies a modified log-Sobolev inequality with constant $\rho(P_{\mathcal{B},\alpha}) \geq \frac{\alpha_{min}}{C_{\rm GBF}}$.*

The main theorem of this section is the following; Theorem 1.10 from the introduction follows as a corollary of this result.

Theorem 5.2. *Let $\eta > 0, b > 0, \Delta \geq 3$ and $\chi \geq 2$. Suppose $G = (V, E)$ is an n-vertex graph of maximum degree Δ and chromatic number χ. Let μ be a Gibbs distribution of a totally-connected spin system on G. Let $\mathcal{B} := \{B_1, \dots, B_K\}$ be any collection of blocks such that $V = \cup_i B_i$, and let α be a distribution over $\mathcal{B}$. If μ is η-spectrally independent and b-marginally bounded, then there exists a universal constant $C > 1$ such that a general heat-bath block dynamics $P_{\mathcal{B},\alpha}$ w.r.t. $(\mathcal{B}, \alpha)$ has modified log-Sobolev constant:*

$$\rho(P_{\mathcal{B},\alpha}) = \Omega\left(\frac{\alpha_{min} \cdot b^{6\kappa}}{\chi \cdot (C\Delta(\eta+1)^5 \log n)^{\kappa \cdot}}\right),$$

where $\kappa = 2 + \lceil \frac{2\eta}{b} \rceil$, and

$$T_{mix}(P_{\mathcal{B},\alpha}) = O\left(\frac{\chi}{\alpha_{min}} \cdot b^{-6\kappa} \cdot \left(C(\eta+1)^5 \Delta \log n\right)^\kappa \cdot \log n\right).$$

Theorem 5.2 follows from the bounds for $C_{\rm KPF}$ in Theorem 3.3 and the following lemma from [BCC⁺22] that relates k-partite factorization with the general block factorization.

Lemma 5.3 (Lemma 3.4, [BCC⁺22])**.** *Suppose the Gibbs distribution μ of a spin system on a graph G satisfies k-partite factorization of entropy with constant $C_{\rm KPF}$. Then μ satisfies GBF with constant $k \cdot C_{\rm KPF}$.*

Proof of Theorem 5.2. The lower bounds for the entropy decay rate and MLSI constant follow from Theorem 3.3, Lemma 5.3 and Proposition 5.1, and by (5) we obtain the desired upper bound for mixing time. □

We also obtain the following corollary for the ferromagnetic Ising and Potts model.

Corollary 5.4. *Let $\delta \in (0, 1)$ and $\Delta \geq 3$. For the Ising model with $\beta \in [(1-\delta)\bar{\beta}_u(\Delta), (1-\delta)\beta_u(\Delta)]$ on any graph G of maximum degree Δ and chromatic number χ, or the ferromagnetic q-state Potts model with $q \geq 2$ and $0 < \beta \leq \frac{2(1-\delta)}{\Delta}$ on the same graph,*

$$T_{mix}(P_{\mathcal{B},\alpha}) = O\left(\frac{\chi}{\alpha_{min}}\right) \cdot O\left(\frac{\Delta}{\delta}\right)^{2+O(1/\delta)} \cdot (\log n)^{3+O(1/\delta)}.$$

Proof of Corollary 5.4. We have shown in the proof of Corollary 3.16 that, for the ferromagnetic q-state Potts model when β is such that $0 < \beta \leq \frac{2(1-\delta)}{\Delta}$, then $b = O(1)$ and $\eta = O(1/\delta)$. For the Ising model, we achieve the same bound by Proposition 4.6. Now $\kappa = 2 + \lceil \frac{2\eta}{b} \rceil = 2 + O(1/\delta)$, and the mixing time bound follows from Theorem 5.2. □

References

[AJK⁺22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: Optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 1418–1430, New York, NY, USA, 2022. Association for Computing Machinery.

[Ale98] Kenneth S. Alexander. On weak mixing in lattice models. *Probab. Theory Relat. Fields*, 110(441-471), 1998.

[ALOG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020.

[AN05] Dimitris Achlioptas and Assaf Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162(3):1335–1351, 2005.

[BCC⁺22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Stefankovic, and Eric Vigoda. On mixing of Markov chains: coupling, spectral independence, and entropy factorization. *Electronic Journal of Probability*, 27:1 – 42, 2022.

[BCP⁺22] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen–Wang dynamics on $\mathbb{Z}^d$. *The Annals of Applied Probability*, 32(2):1018 – 1057, 2022.

[BCSV19] Antonio Blanca, Pietro Caputo, Alistair Sinclair, and Eric Vigoda. Spatial mixing and nonlocal markov chains. *Random Structures & Algorithms*, 55(3):584–614, 2019.

[BCSV23] Antonio Blanca, Zongchen Chen, Daniel Stefankovic, and Eric Vigoda. The Swendsen–Wang dynamics on trees. *Random Structures & Algorithms*, 2023.

[BCT12] Christian Borgs, Jennifer T. Chayes, and Prasad Tetali. Tight bounds for mixing of the Swendsen-Wang algorithm at the Potts transition point. *Probab. Theory Relat. Fields*, 152(3-4):509–557, 2012.

[BCV20] Antonio Blanca, Zongchen Chen, and Eric Vigoda. Swendsen-wang dynamics for general graphs in the tree uniqueness region. *Random Structures & Algorithms*, 56(2):373–400, 2020.

[BDC12] Vincent Beffara and Hugo Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probab. Theory Relat. Fields*, 153(511-542), 2012.

[BGP16] Magnus Bordewich, Catherine Greenhill, and Viresh Patel. Mixing of the glauber dynamics for the ferromagnetic potts model. *Random Structures & Algorithms.*, 48(1):21–52, January 2016.

[BS15] Antonio Blanca and Alistair Sinclair. Dynamics for the mean-field random-cluster model. In *Proceedings of APPROX/RANDOM*, 2015.

[BT03] Sergey Bobkov and Prasad Tetali. Modified log-sobolev inequalities, mixing and hypercontractivity. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 287–296, New York, NY, USA, 2003. Association for Computing Machinery.

[Ces01] Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probability Theory and Related Fields*, 120:569–584, 2001.

[CFYZ22a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 588–599. IEEE, 2022.

[CFYZ22b] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of glauber dynamics via spectral independence for all degrees. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 137–148, 2022.

[CGSV21] Zongchen Chen, Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557, 2021.

[CLMM23] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications, 2023.

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318. IEEE Computer Society, 2020.

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 1537–1550, New York, NY, USA, 2021. Association for Computing Machinery.

[CLY23] Xiaoyu Chen, Jingcheng Liu, and Yitong Yin. Uniqueness and rapid mixing in the bipartite hardcore model, 2023.

[CMT14] Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature, 2014.

[COGG+23] Amin Coja-Oghlan, Andreas Galanis, Leslie Ann Goldberg, Jean Bernoulli Ravelomanana, Daniel Štefankovič, and Eric Vigoda. Metastability of the Potts ferromagnet on random regular graphs. *Communications in Mathematical Physics*, 2023.

[CP21] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388(2):793–818, oct 2021.

[DGJ06a] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Dobrushin conditions and systematic scan. In Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 327–338, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[DGJ06b] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Systematic scan for sampling colorings. *The Annals of Applied Probability*, 16(1):185 – 230, 2006.

[DGJ09] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71 – 107, 2009.

[DSC96] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695 – 750, 1996.

[DSVW04] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24(4):461–479, 2004.

[FGYZ22] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. *ACM Trans. Algorithms*, 18(3), oct 2022.

[Fil91] James Allen Fill. Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process. *The Annals of Applied Probability*, 1(1):62 – 87, 1991.

[FK13] James Allen Fill and Jonas Kahn. Comparison inequalities and fastest-mixing Markov chains. *The Annals of Applied Probability*, 23(5):1778 – 1816, 2013.

[GJ97] Vivek K. Gore and Mark R. Jerrum. The Swendsen-Wang process does not always mix rapidly. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 674–681, New York, NY, USA, 1997. Association for Computing Machinery.

[GJ17] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017*, pages 1818–1827, 2017.

[GKZ18] Heng Guo, Kaan Kara, and Ce Zhang. Layerwise systematic scan: Deep Boltzmann machines and beyond. In *International Conference on Artificial Intelligence and Statistics*, pages 178–187. PMLR, 2018.

[GL18] Reza Gheissari and Eyal Lubetzky. Mixing times of critical two-dimensional Potts models. *Comm. Pure Appl. Math*, 71(5):994–1046, 2018.

[GLP19] Reza Gheissari, Eyal Lubetzky, and Yuval Peres. Exponentially slow mixing in the mean-field Swendsen–Wang dynamics. *Annales de l'Institut Henri Poincare (B)*, 2019. Extended abstract appeared in Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2018), pp. 1981–1988.

[GŠV15] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. In *Proceedings of APPROX/RANDOM*, 2015.

[Hay06] Thomas P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *The 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 39–46, 2006.

[HS07] Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *The Annals of Applied Probability*, 17(3), Jun 2007.

[Hub03] Mark Huber. A bounding chain for Swendsen-Wang. *Random Structures & Algorithms*, 22(1):43–59, 2003.

[JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. Spectral independence, coupling, and the spectral gap of the glauber dynamics. *Information Processing Letters*, 177:106268, 2022.

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, 2006.

[Mar99] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. *Lectures on probability theory and statistics (Saint-Flour, 1997)*, 1717:93–191, 1999.

[Mar19] Katalin Marton. Logarithmic sobolev inequalities in discrete product spaces. *Combinatorics, Probability and Computing*, 28(6):919–935, 2019.

[MOS94] 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.

[MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

[PW13] Yuval Peres and Peter Winkler. Can extra updates delay mixing? *Communications in Mathematical Physics*, 323(3):1007–1016, 2013.

[RST+13] Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. Improved mixing condition on the grid for counting and sampling independent sets. *Probab. Theory Relat. Fields*, 2013.

[SIM93] BARRY SIMON. *The Statistical Mechanics of Lattice Gases, Volume I.* Princeton University Press, 1993.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

[SSSY17] Alistair Sinclair, Piyush Srivastava, Daniel Stefankovic, and Yitong Yin. Spatial mixing and the connective constant: Optimal bounds. *Probab. Theory Relat. Fields*, 168:153–197, 2017.

[SW87] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[Ull14] Mario Ullrich. Rapid mixing of swendsen–wang dynamics in two dimensions. *Dissertationes Mathematicae*, 502:1–64, 2014.

A Proof of the second part of Theorem 3.3

In this appendix, we prove (9) in Theorem 3.3, which begins by extrapolating the proof of Lemma 3.3 in [BCC+22] as Lemma A.1.

Lemma A.1 ([BCC+22]). *Let $\theta \in (0,1]$ and $n \geq \frac{2}{\theta}(\frac{4\eta}{b^2}+1)$. Let $G, \mu, V_1, \ldots, V_k$ be as in the assumption of Theorem 3.3. Let S be a uniformly generated block of vertices of size $\lceil \theta n \rceil$, and let $S_1, \ldots, S_m$ be the connected components of S. Recall that $C_S(v)$ denotes the unique connected component S_i in S that contains v if such a component exists, otherwise set it to be the empty set. Suppose further that for $S_i \subseteq S$, $\Gamma(S_i)$ takes the* *minimum value such that the following inequality holds for an arbitrary pinning $\tau \in \Omega_{S_i^c}$ and any function $g : \Omega^{\tau}_{S_i} \to \mathbb{R}_{\geq 0}$:*

$$\mathrm{Ent}^{\tau}_{S_i}(g) \leq \Gamma(S_i) \sum_{j=1}^{k} \mathrm{E}_{\xi \sim \mu^{\tau}_{S_i \setminus V_j}} \left[\mathrm{Ent}^{\xi \cup \tau}_{V_j \cap S_i}(g^{\xi}_{S_i \cap V_j}) \right].$$

Then,

$$\mathrm{Ent}_{\mu}(f) \leq \frac{C_{\mathrm{UBF}}}{\theta} \sum_{j=1}^{k} \mathrm{E}_{\tau \sim \mu} \left[\mathrm{Ent}^{\tau}_{V_j}(f) \right] \cdot G_j, \tag{38}$$

where

$$G_j := \max_{W \subset V_j} \max_{v \in W} \mathrm{E}_S \left[\Gamma(C_S(v)) \mid V_j \cap S = W \right]$$

and the expectation E_S is taken over the uniform generation of S.

Proof of (9) *in Theorem 3.3.* In the same way that we prove (8), if $\Delta^2 > \frac{b^4 n}{10e(4\eta + b^2)}$ then it follows from Lemma 3.7 and Lemma 3.8 that

$$C_{\mathrm{KPF}} \leq \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^{\kappa}} \cdot \left(\frac{10e(4\eta + b^2)}{b^2} \right)^{\kappa} \cdot \Delta^{2\kappa} \leq \frac{(240e)^{4\kappa} \cdot (\lceil \eta \rceil + 1)^{5\kappa} \cdot \Delta^{2\kappa}}{b^{6\kappa}}.$$

Now we assume $\Delta^2 \leq \frac{b^2 n}{10e(4\eta + b^2)}$. Take $\theta = \frac{1}{5e\Delta^2} \geq \frac{2(4\eta + b^2)}{b^2 n} = \frac{2}{n} \cdot (\frac{4\eta}{b^2} + 1)$. Theorem 3.2 implies that

$$C_{\mathrm{UBF}} = \left(\frac{e}{\theta} \right)^{\lceil \frac{2\eta}{b} \rceil} = \left(5e^2 \Delta^2 \right)^{\lceil \frac{2\eta}{b} \rceil}. \tag{39}$$

Given Lemma A.1, to show (9) it remains to provide an upper bound G_j for each j. There are two main steps for proving this bound. First, we upper bound G_j in terms of the size of connected components in S. Under the assumptions of Theorem 3.3, μ is η-spectrally independent and b-marginally bounded. These properties by definition preserve under any pinning. In particular, for any $S_i \subseteq S$ and an arbitrary pinning $\tau \in \Omega_{V \setminus S_i}$, $\mu^{\tau}_{S_i}$ is still η-spectrally independent and b-marginally bounded. Thus, Lemma 3.7 and Lemma 3.8 imply that

$$\Gamma(S_i) \leq \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^{\kappa}} \cdot |S_i|^{\kappa},$$

and letting $\tilde{b} := \frac{3(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^{\kappa}}$ we have

$$G_j \leq \tilde{b} \max_{W \subset V_j} \max_{v \in W} \mathrm{E}_S \left[|C_S(v)|^{\kappa} \mid V_j \cap S = W \right]. \tag{40}$$

The second part of this proof analyzes the conditional expectation term above on the right-hand side of (40). We fix $v \in V$ (and hence fix V_j) and fix a feasible W such that $v \in W \subseteq V_j$ and $|W| \leq \lceil \theta n \rceil$. We say a set $T \subseteq V \setminus V_j$ is *W-connected* if $T \cup W$ is connected in G, and we denote by $S'(v)$ the unique W-connected vertex-set in S that is adjacent to v, if such set exists, otherwise an empty set. Clearly if $S'(v) = \emptyset$, then $C_S(v) = \{v\}$. Suppose $S'(v) \neq \emptyset$. Observe that $C_S(v) = S'(v) \cup (C_S(v) \cap W)$. Since $(C_S(v) \cap W)$ must be adjacent to $S'(v)$ if $S'(v) \neq \emptyset$, $|C_S(v) \cap W| \leq \Delta \cdot |S'(v)|$. Hence, $|C_S(v)| \leq (\Delta + 1)|S'(v)|$.

Furthermore, let $G_2 := (V, E \cup E_2)$, where E_2 is the set of pairs of vertices that are of distance at most 2 in G. Note that the degree of any vertex in G_2 is at most Δ^2. Let $C_{S_2}(v)$ be the unique connected component in $G_2[S]$ that contains v. Notice that the set $S'(v)$ is always a subset of $C_{S_2}(v)$, regardless of the specific set W we choose to fix. Hence, for any x,

$$\Pr_S \left[|C_S(v)| \geq x \mid V_j \cap S = W \right] \leq \Pr_S \left[|S'(v)| \geq \frac{x}{\Delta + 1} \mid V_j \cap S = W \right] \leq \Pr_S \left[|C_{S_2}(v)| \geq \frac{x}{\Delta + 1} \right].$$

Now we apply Lemma 3.6 to estimate the last probability. For $\theta < \frac{1}{4e\Delta^2}$,

$$\begin{aligned}
\Pr_S\left[|C_{S_2}(v)| \geq \frac{x}{\Delta+1}\right] &\leq \frac{\lceil \theta n \rceil}{n} \sum_{k=0}^{\infty} (2e\Delta^2\theta)^{\lfloor \frac{x}{\Delta+1} \rfloor + k - 1} \\
&\leq \frac{1}{2e\Delta^2}\left(\frac{1}{2}\right)^{\lfloor \frac{x}{\Delta+1} \rfloor} \cdot \sum_{k=0}^{\infty} \frac{1}{2} \\
&\leq \frac{1}{\Delta^2} \cdot 2^{-\frac{x}{\Delta+1}}.
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\mathrm{E}_S\left[|C_S(v)|^\kappa \mid V_j \cap S = W\right] &\leq \sum_{x=1}^{n} x^\kappa \Pr_S\left[|C_S(v)| \geq x \mid V_j \cap S = W\right] \\
&\leq \sum_{x=1}^{n} x^\kappa \Pr_S\left[|C_{S_2}(v)| \geq \frac{x}{\Delta+1}\right] \\
&\leq \sum_{x=1}^{n} x^\kappa \cdot \frac{1}{\Delta^2} \cdot 2^{-\frac{x}{\Delta+1}} \\
&\leq 4\Delta^{2\kappa}.
\end{aligned}$$

Therefore, $G_j \leq 4\tilde{b}\Delta^{2\kappa}$. This bound on G_j together with (38) and (39) implies

$$C_{\mathrm{KPF}} \leq 4\tilde{b}\Delta^{2\kappa} \cdot \left(5e^2\Delta^2\right)^\kappa = \frac{12(\lceil 2\eta \rceil + 2)^{4\kappa}}{(2b^4)^\kappa} \cdot \left(5e^2\Delta^2\right)^\kappa \cdot \Delta^{2\kappa},$$

concluding the proof. □

B Additional proofs

Proof of Lemma 3.7. Let $\eta_0, \eta_1, \ldots, \eta_{n-2}$ be a sequence of reals. We say a distribution μ is $(\eta_0, \eta_1, \ldots, \eta_{n-2})$-spectrally independent if for every $0 \leq k \leq n-2$, any $\Lambda \subseteq V$ of size k and any pinning τ on Λ, $\lambda_1(\Psi_\mu^\tau) \leq \eta_k$. Theorem 6 and 8 from [CGSV21][1] state that if μ is $(\eta_0, \eta_1, \ldots, \eta_{n-2})$-spectrally independent, then the spectral gap of the Glauber dynamics is at least

$$\frac{1}{n} \prod_{k=0}^{n-2} \left(1 - \frac{\eta_k}{n-k-1}\right). \tag{41}$$

We complete the proof by establishing suitable bounds for each η_k. Per Definition 1.1, we have $\eta_k \leq \eta$ for all $k \in [0, n-2]$. In addition, we will show that

$$\eta_k \leq (n-k-1) \cdot \left(1 - \frac{2b^4}{(n-k)^4}\right). \tag{42}$$

[1]Originally these theorems are given for coloring, but their proofs naturally extend to general totally-connected distributions.

As such, we will have that $\eta_k \le \min\{\eta, (n-k-1)\cdot(1-\frac{2b^4}{(n-k)^4})\}$, and we would finish the proof of Lemma 3.7 by plugging these bounds for η_k into (41):

$$\frac{1}{n}\prod_{k=0}^{n-2}\left(1-\frac{\eta_k}{n-k-1}\right) \ge \frac{1}{n}\prod_{k=0}^{n-2}\left(1-\min\left\{\frac{\eta}{n-k-1}, 1-\frac{2b^4}{(n-k)^4}\right\}\right) = \frac{1}{n}\prod_{k=1}^{n-1}\left(1-\min\left\{\frac{\eta}{k}, 1-\frac{2b^4}{(k+1)^4}\right\}\right)$$
$$\ge \frac{1}{n}\left(\prod_{k=1}^{\lceil 2\eta\rceil+1}\frac{2b^4}{(k+1)^4}\right)\left(\prod_{k=\lceil 2\eta\rceil+2}^{n-1}\left(1-\frac{\eta}{k}\right)\right) \ge \frac{1}{n}\left(\frac{2b^4}{(\lceil 2\eta\rceil+2)^4}\right)^{\lceil 2\eta\rceil+1}\cdot\exp\left(-\sum_{k=\lceil 2\eta\rceil+2}^{n-1}\frac{2\eta}{k}\right)$$
$$\ge \frac{1}{n}\left(\frac{2b^4}{(\lceil 2\eta\rceil+2)^4}\right)^{\lceil 2\eta\rceil+1}\cdot\exp\left(-2\eta\ln n\right) \ge \left(\frac{2b^4}{(\lceil 2\eta\rceil+2)^4}\cdot\frac{1}{n}\right)^{1+\lceil 2\eta\rceil}.$$

Now we provide a proof for (42). Let τ be a pinning on Λ with $|\Lambda| = k$, and let $U = V \setminus \Lambda$. Theorem 8 of [CGSV21] shows that

$$\lambda_1(\Psi_U^\tau) = (n-k-1)\cdot\lambda_2(\hat{P}_\tau),$$

where $\hat{P}_\tau$ denotes the transition matrix of the *local random walk* on $\mathcal{P}^\tau := \{(u,s) : u \notin \Lambda, s \in \Omega_u^\tau\}$ whose entries are given by $\hat{P}_\tau((u,a),(v,b)) := \frac{\mathbb{1}[u\neq v]}{n-k-1}\cdot\mu_{U\setminus\{u\}}^{\tau\cup(u,a)}(\sigma_v = b)$. Let π^τ be a distribution on $\mathcal{P}^\tau$ given by $\pi^\tau(u,s) = \frac{1}{n-k}\cdot\mu^\tau(\sigma_u = s)$. It is straightforward to verify that $\hat{P}_\tau$ is reversible with respect to π^τ. By the standard relationship between conductance and the eigenvalue of a reversible transition matrix in [SJ89], we have

$$1-\lambda_2(\hat{P}_\tau) \ge \frac{\Phi^2}{2},$$

where

$$\Phi := \min_{S\subseteq\mathcal{P}^\tau: S\neq\emptyset, \pi^\tau(S)\le 1/2}\Phi_S, \text{ and } \quad \Phi_S := \frac{1}{\pi^\tau(S)}\sum_{x\in S}\sum_{y\notin S}\pi^\tau(x)\hat{P}_\tau(x,y).$$

As μ is totally-connected, for any $S \subseteq \mathcal{P}^\tau$ such that $S \neq \emptyset$ and $\pi^\tau(S) \le 1/2$, there exist $x \in S$ and $y \notin S$ such that $\hat{P}_\tau(x,y) > 0$. Also, since μ is b-marginally bounded, we have $\pi^\tau(x) \ge b/(n-k)$ and $\hat{P}_\tau(x,y) \ge b/(n-k-1)$. Hence,

$$\Phi \ge 2\min_{S\subseteq\mathcal{P}^\tau,\pi^\tau(S)\le 1/2}\ \min_{x\in S, y\notin S:\hat{P}_\tau(x,y)>0}\pi^\tau(x)\hat{P}_\tau(x,y) \ge 2\cdot\frac{b}{n-k}\cdot\frac{b}{n-k-1} \ge \frac{2b^2}{(n-k)^2}.$$

It follows that

$$\frac{\lambda_1(\Psi_U^\tau)}{n-k-1} = 1-\mathrm{GAP}(\hat{P}_\tau) \le 1-\frac{\Phi^2}{2} \le 1-\frac{2b^4}{(n-k)^4},$$

which establishes (42). □

Proof of Lemma 3.8. We say that μ satisfies the log-Sobolev inequality with constant ρ_1 if for all functions $f : \Omega \to \mathbb{R}_{\ge 0}$,

$$\rho_1\mathrm{Ent}_\mu(f) \le \frac{1}{n}\sum_{v\in V}\mathrm{E}_{\tau\sim\mu_{V\setminus\{v\}}}\left[Var_v^\tau(\sqrt{f^\tau})\right].$$

Recall that C_{AT} is the least constant such that for all functions $f : \Omega \to \mathbb{R}_{\ge 0}$,

$$\mathrm{Ent}_\mu(f) \le C_{\mathrm{AT}}\sum_{v\in V}\mathrm{E}_{\tau\sim\mu_{V\setminus\{v\}}}\left[\mathrm{Ent}_v^\tau(f^\tau)\right].$$

Proposition 1.1 from [CMT14] implies that

$$C_{AT} \le \frac{1}{\rho_1 n} \tag{43}$$

Moreover, [DSC96] shows that

$$\frac{1 - 2\mu_{min}}{\log(1/\mu_{min} - 1)} \cdot \gamma \leq \rho_1. \tag{44}$$

If $\mu_{min} > 1/3$, then μ is a trivial distribution and $C_{AT} \leq 1$. Thus, we may assume that $\mu_{min} \leq 1/3$. Since μ is b-marginally bounded, we have

$$\frac{1 - 2\mu_{min}}{\log(1/\mu_{min} - 1)} \geq \frac{1}{3\log(1/\mu_{min})} \geq \frac{1}{3n\log(b^{-1})}. \tag{45}$$

It follows from (43), (44) and (45) that

$$C_{AT} \leq \frac{3\log(b^{-1})}{\gamma}. \tag{46}$$

Observe that by Corollary 3.12, if $v \in B$, then

$$\mathrm{E}_{\tau \sim \mu_{V\setminus\{v\}}}\left[\mathrm{Ent}_v^\tau(f^\tau)\right] \leq \mathrm{E}_{\tau \sim \mu_{V\setminus B}}\left[\mathrm{Ent}_B^\tau(f^\tau)\right].$$

Hence, given k disjoint independent sets $U_1, \ldots, U_k$ of V such that $\bigcup_{i=1}^k U_i = V$, we have

$$\sum_{v\in V} \mathrm{E}_{\tau \sim \mu_{V\setminus\{v\}}}\left[\mathrm{Ent}_v^\tau(f^\tau)\right] = \sum_{j=1}^k \sum_{v\in U_j} \mathrm{E}_{\tau \sim \mu_{V\setminus\{v\}}}\left[\mathrm{Ent}_v^\tau(f^\tau)\right] \leq n \sum_{j=1}^k \mathrm{E}_{\tau \sim \mu_{V\setminus U_j}}\left[\mathrm{Ent}_v^\tau(f^\tau)\right].$$

Equivalently, we obtain that

$$C_{KPF} \leq n \cdot C_{AT}. \tag{47}$$

By (46) and (47), we establish the lemma. □