On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization*

Antonio Blanca† Pietro Caputo‡ Zongchen Chen§ Daniel Parisi†
Daniel Štefankovič¶ Eric Vigoda‖

Abstract

For general spin systems, we prove that a contractive coupling for an arbitrary local Markov chain implies optimal bounds on the mixing time and the modified log-Sobolev constant for a large class of Markov chains including the Glauber dynamics, arbitrary heat-bath block dynamics, and the Swendsen-Wang dynamics. This reveals a novel connection between probabilistic techniques for bounding the convergence to stationarity and analytic tools for analyzing the decay of relative entropy. As a corollary of our general results, we obtain $O(n \log n)$ mixing time and $\Omega(1/n)$ modified log-Sobolev constant of the Glauber dynamics for sampling random q-colorings of an n-vertex graph with constant maximum degree Δ when $q > (11/6 - \epsilon_0)\Delta$ for some fixed $\epsilon_0 > 0$. We also obtain $O(\log n)$ mixing time and $\Omega(1)$ modified log-Sobolev constant of the Swendsen-Wang dynamics for the ferromagnetic Ising model on an n-vertex graph of constant maximum degree when the parameters of the system lie in the tree uniqueness region. At the heart of our results are new techniques for establishing spectral independence of the spin system and block factorization of the relative entropy. On one hand we prove that a contractive coupling of any local Markov chain implies spectral independence of the Gibbs distribution. On the other hand we show that spectral independence implies factorization of entropy for arbitrary blocks, establishing optimal bounds on the modified log-Sobolev constant of the corresponding block dynamics.

1 Introduction

Spectral independence is a powerful new approach for proving fast convergence of Markov chain Monte Carlo (MCMC) algorithms. The technique was introduced by Anari, Liu, and Oveis Gharan [2] to establish rapid mixing of the Glauber dynamics by utilizing high-dimensional expanders. For a spin system defined on a graph $G = (V, E)$, the Glauber dynamics is the simple single-site update Markov chain which updates the spin at a randomly chosen vertex in each step. The mixing time is the number of steps to reach close to the stationary distribution.

Our paper addresses two broad questions. First, what are the implications of spectral independence? In particular, does it imply fast convergence for other Markov chains beyond the simple Glauber dynamics? We prove that it does: we show that spectral independence implies optimal mixing time bounds and modified log-Sobolev constants for a broad class of chains, including all possible heat-bath block dynamics and the Swendsen-Wang dynamics. Our proof utilizes recent work on entropy factorization [12, 5].

Our second question is when does spectral independence hold, and how does it relate to traditional proof approaches, such as coupling techniques? Here again we prove a general result, showing that a contractive coupling for any *local* Markov chain implies spectral independence. This immediately yields stronger than state of the art mixing time bounds for a variety of chains. In addition, it provides an intriguing conceptual connection between the coupling method and modified log-Sobolev inequalities as we describe below.

*The full version of the paper can be accessed at https://arxiv.org/abs/2103.07459

†Department of Computer Science and Engineering, Penn State, University Park, PA 16801, USA. Email: ablanca@cse.psu.edu. Research supported in part by NSF grant CCF-1850443.

‡Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy. Email: {pietro.caputo,daniel.parisi}@uniroma3.it.

§Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: zongchen@mit.edu

¶Department of Computer Science, University of Rochester, Rochester, NY 14627, USA. Email: stefanko@cs.rochester.edu. Research supported in part by NSF grant CCF-2007287.

‖Department of Computer Science, University of California, Santa Barbara, CA, 93106, USA. Email: vigoda@ucsb.edu. Research supported in part by NSF grant CCF-2007022.

There are two broad approaches for establishing fast convergence of MCMC algorithms: probabilistic or analytic techniques. Probabilistic techniques primarily utilize the coupling method; a popular example is the path coupling method which has become a fundamental tool in theoretical computer science [10]. In contrast, analytic techniques establish decay to equilibrium by means of functional inequalities such as Poincaré or log-Sobolev inequalities, which correspond to decay of variance and relative entropy respectively. In particular, the so-called modified log-Sobolev inequality is often a powerful analytic tool in establishing tight bounds on the mixing time, while the weaker Poincaré inequality provides control on the spectral gap; see, e.g., [19, 33, 7].

These two approaches—probabilistic or analytic—appeared disparate. While coupling techniques have been used to prove Poincaré inequalities, there are no clear relations between the probabilistic approach and log-Sobolev inequalities. We establish a strong connection by proving that coupling inequalities in the form of bounds on the Ollivier-Ricci curvature of the Markov chain imply entropy decay, and hence the associated modified log-Sobolev inequality holds; see Section 2 for definitions. In the specific context of spin systems on bounded-degree graphs, this settles a remarkable (and more general) conjecture of Peres and Tetali (see Conjecture 3.1 in [22] and Remark 1.1).

Our technical contributions apply in the general setting of q-state spin systems. This is a convenient setting to capture a wide family of distributions defined on graphs, including the equilibrium distribution of undirected graphical models. We now introduce some relevant notation and refer to Section 2 for a formal definition of general spin systems. Let $G = (V, E)$ be an n-vertex graph, and let Δ denote the maximum degree of G. For integer $q \geq 2$ the state space of the model is the set $\Omega = \{\sigma \in [q]^V : \mu(\sigma) > 0\}$ of assignments with positive weight in the Gibbs distribution μ.

Canonical examples of a spin system include the Ising model (with $q = 2$ spin values) and the Potts model (with $q \geq 3$); in these models, for an inverse temperature parameter β, a configuration $\sigma \in \Omega$ has probability $\mu(\sigma) \propto \exp(\beta M(\sigma))$ where $M(\sigma)$ is the number of edges of G which are monochromatic in σ. The Ising/Potts model is ferromagnetic when $\beta > 0$ and antiferromagnetic when $\beta < 0$. The hard-core model is another spin system defined on the set of independent sets of G weighted by a parameter $\lambda > 0$; each independent set σ has probability proportional to $\lambda^{|\sigma|}$ in the Gibbs distribution. The q-colorings model, where the Gibbs distribution is uniform over the collection of proper vertex q-colorings of G, is also a classical spin system.

The Glauber dynamics is the simplest MCMC approach for sampling from the Gibbs distribution μ. The transitions of the Markov chain (X_t) update the spin at a randomly chosen vertex in each step. From $X_t \in \Omega$, we choose a random vertex x, set $X_{t+1}(y) = X_t(y)$ for all $y \neq x$, and the spin $X_{t+1}(x)$ is chosen from the marginal distribution at x conditional on the current spins on $N(x)$, the neighborhood of x. The mixing time is the number of steps, from the worst initial state, to get close to the stationary distribution; see Section 2.

Our results apply more broadly to the general class of heat-bath block dynamics. Let $\mathcal{B} = \{B_1, \ldots, B_\ell\}$ be any collection of sets (or blocks) such that $V = \cup_i B_i$ and let $\alpha = (\alpha_B)_{B \in \mathcal{B}}$ be a probability distribution on $\mathcal{B}$. A step of the heat-bath block dynamics operates by choosing a block $B \in \mathcal{B}$ with probability α_B and updating the configuration in B with a sample from the Gibbs distribution conditional on the configuration on $V \setminus B$. Note that the Glauber dynamics corresponds to setting the blocks to individual vertices with uniform weights, and for a bipartite graph, the even-odd chain (also known as the alternating scan dynamics) corresponds to uniform weighting for two blocks corresponding to the two parts. By extending the weight to $\alpha_B = 0$ if $B \notin \mathcal{B}$ we think of α as a distribution over all subsets of V and speak of the α-weighted heat-bath block dynamics.

Given α, define the minimum "coverage probability" of a vertex by

$$\delta = \delta(\alpha) = \min_{x \in V} \sum_{B : B \ni x} \alpha_B. \tag{1.1}$$

We say that the block dynamics have *optimal mixing* when there exists a constant C such that for all weights α the mixing time of the α-weighted heat-bath block dynamics is at most $C\delta(\alpha)^{-1} \log n$. Similarly, we say that the block dynamics have *optimal entropy decay* if the modified log-Sobolev constant of the α-weighted heat-bath block dynamics is at least $\delta(\alpha)/C$. Note that the constant C may depend on the parameters defining the spin system and on the maximum degree Δ, but it does not depend on n and it is independent of the choice of weights α. In this generality, these bounds are optimal up to the value of the constant C. Indeed, for the Glauber dynamics we have $\delta(\alpha) = 1/n$ and the mixing time matches the $\Omega(n \log n)$ lower bound established by Hayes and Sinclair [28] for bounded-degree graphs. Moreover, by restricting to test functions of a single spin it is not hard to check that the spectral gap of the α-weighted block dynamics is always at most $\delta(\alpha)$, and therefore the lower bound $\delta(\alpha)/C$

on the modified log-Sobolev constant of the block dynamics is optimal up to the multiplicative constant $1/C$; see e.g. [7] for standard relations between spectral gap and modified log-Sobolev constant.

1.1 Applications We begin with a few examples of applications of our results. We then delve into our general technical contributions in subsequent subsections. We note that all of these applications follow immediately from previous coupling proofs together with our new technical contributions.

For q-colorings of graphs with maximum degree Δ, Jerrum [30] proved that the Glauber dynamics has $O(n \log n)$ mixing time when $q > 2\Delta$. Jerrum's result was improved to $q > \frac{11}{6}\Delta$ in [44] and further improved to $q > (\frac{11}{6} - \epsilon_0)\Delta$ for some small $\epsilon_0 \approx 10^{-5} > 0$ by Chen et al. [14] by analyzing a Markov chain referred to as the flip dynamics; this implied $O(n^2)$ mixing time of the Glauber dynamics. We obtain $O(n \log n)$ mixing time of the Glauber dynamics, which is asymptotically optimal [28], and also obtain optimal bounds on the log-Sobolev and modified log-Sobolev constants.

THEOREM 1.1. *For q-colorings on an n-vertex graph of maximum degree Δ, when $q > (\frac{11}{6} - \epsilon_0)\Delta$, where $\epsilon_0 \approx 10^{-5} > 0$ is a fixed constant, the Glauber dynamics has mixing time $O(n \log n)$ and log-Sobolev and modified log-Sobolev constants $\Omega(1/n)$. More generally, under these assumptions all block dynamics have optimal mixing and optimal entropy decay.*

For the ferromagnetic Ising model, Mossel and Sly [37] established optimal mixing time bounds of $O(n \log n)$ for the Glauber dynamics on any graph of maximum degree Δ in the tree uniqueness region; that is, for all $\beta < \beta_c(\Delta)$, where $\beta_c(\Delta) := \ln(\frac{\Delta}{\Delta-2})$ is the threshold of the uniqueness/non-uniqueness phase transition on the Δ-regular tree. Our general results allow us to extend this to arbitrary heat-bath block dynamics and to the Swendsen-Wang dynamics [42]. The latter is a particularly interesting Markov chain which utilizes the random-cluster representation of the ferromagnetic Potts model to perform global updates in a single step. This non-local nature makes tight analysis of the Swendsen-Wang dynamics challenging. In [6], it was shown that the mixing time of Swendsen-Wang dynamics on any graph of maximum degree Δ in the tree uniqueness region is $O(n)$. Our general results imply a bound of $O(\log n)$ on the mixing time of the Swendsen-Wang dynamics and a bound of $\Omega(1)$ on the corresponding modified log-Sobolev constant in the same tree uniqueness region. As shown in [5] for the special case of the d-dimensional integer lattice $\mathbb{Z}^d$, these estimates are optimal up to a multiplicative constant. Our results also yield new optimal bounds on the log-Sobolev and modified log-Sobolev constants for the Glauber dynamics in the same setting.

We also obtain improved results for the ferromagnetic Potts model. Unlike the Ising model, for the ferromagnetic Potts model known rapid mixing results for the Glauber dynamics do not reach the tree uniqueness threshold. The best known results [27, 43, 8] imply that the Glauber dynamics mixes in $O(n \log n)$ steps when $\beta < \beta_0$ where $\beta_0 = \max\{\frac{2}{\Delta}, \frac{1}{\Delta}\ln(\frac{q-1}{\Delta})\}$. In addition, [8] showed poly(n) mixing of the Glauber dynamics for $\beta < \beta_1$ where $\beta_1 = (1 - o(1))\frac{\ln q}{\Delta-1}$, the $o(1)$ term tends to 0 as $q \to \infty$; see Remark 3.5 for more details. These results yield polynomial mixing time bounds for the Swendsen-Wang dynamics in the corresponding regimes of β. Note the critical point for the uniqueness threshold on the tree was established by Häggström [26] and it behaves as $\beta_u = \frac{\ln q}{\Delta-1} + O(1)$; see [8]. In both regimes, we prove optimal bounds for the mixing time and (modified) log-Sobolev constant of the Glauber dynamics and also for the Swendsen-Wang dynamics.

THEOREM 1.2. *For the* ferromagnetic Ising model *with $\beta < \beta_c(\Delta)$ on any n-vertex graph of maximum degree $\Delta \geq 3$, all heat-bath block dynamics have optimal mixing and optimal entropy decay, and the Swendsen-Wang dynamics has optimal mixing time $O(\log n)$ and optimal modified log-Sobolev constant $\Omega(1)$. For the* ferromagnetic Potts model *the same results hold when $\beta < \max\{\beta_0, \beta_1\}$.*

1.2 Spectral independence definitions A central concept in our work is *spectral independence*, which was introduced by Anari, Liu and Oveis Gharan [2] to establish polynomial mixing time bounds for the Glauber dynamics. To formally define spectral independence it will be important to consider the effect of *pinnings* which can informally be viewed as boundary conditions. For $U \subset V$, let $\Omega_U = \{\tau \in [q]^U : \exists \sigma \in \Omega, \sigma_U = \tau\}$ denote the set of assignments to U with valid extensions on the remaining vertices. In particular, Ω_x denotes the set of all valid spin assignments for the vertex x under μ. A pinning is a fixed assignment τ on some $U \subset V$ where $\tau \in \Omega_U$. We write μ^τ for the Gibbs measure $\mu(\cdot \,|\, \sigma_U = \tau)$ obtained by conditioning on the given τ. In the presence of a pinning τ on $U \subset V$, the definition of the Glauber dynamics remains the same with the assignment

τ on U fixed (see Remark 3.1 for a definition). Let $\mathcal{T} = \cup_{U \subset V} \Omega_U$ denote the collection of all pinnings, and $\mathcal{X} = \{(x, a) : x \in V, a \in \Omega_x\}$ for the set of all feasible vertex-spin pairs.

The spectral independence approach considers the following matrices which capture the pairwise influence of vertices. For a pair of vertices x, y and a pair of spins a, a', it is the influence of the spin a at x on the marginal probability of a' at y.

DEFINITION 1.1. (ALO INFLUENCE MATRIX) *The* ALO influence matrix $J \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$ *is defined by* $J(x, a; x, a') = 0$ *and*

$$J(x, a; y, a') = \mu(\sigma_y = a' \mid \sigma_x = a) - \mu(\sigma_y = a') \quad \text{for } x \neq y.$$

Moreover, for a pinning $\tau \in \mathcal{T}$, J^τ *denotes the influence matrix with respect to the conditional measure* μ^τ.

Note that [2] defined the influence matrix only for $q = 2$ in a slightly different form and the definition was later generalized to all $q \geq 2$ by two independent works [15, 23] in different ways. In this paper we use the definition from [15] which is more suitable for our applications in Section 3 for establishing spectral independence, but we could also work with the definition from [23] with some additional effort. Since J is self-adjoint the eigenvalues of J are real. Let $\lambda_1(J) \geq 0$ denote its largest eigenvalue (the eigenvalue zero always exists since all row sums of J vanish).

DEFINITION 1.2. (SPECTRAL INDEPENDENCE) *We say that a spin system is* η-spectrally independent *if for all pinnings* $\tau \in \mathcal{T}$ *we have* $\lambda_1(J^\tau) \leq \eta$.

There is one additional property of the Gibbs distribution that will be relevant to us; namely, that the marginal probability for any vertex is lower bounded by a constant b. This property is typically trivial to satisfy for some constant $b = b(\Delta) > 0$. We write Ω_x^τ for the set of spin values that are allowed at x in the presence of the pinning τ.

DEFINITION 1.3. (MARGINAL BOUNDEDNESS) *We say that the spin system is* b-marginally bounded *if for all pinnings* τ, *all* $x \in V$, *all* $a \in \Omega_x^\tau$ *we have* $\mu^\tau(\sigma_x = a) \geq b$.

1.3 Consequences of spectral independence The spectral independence approach has been quite powerful as it led to rapid mixing results for the hard-core model in the tree uniqueness region [2], for any 2-spin antiferromagnetic spin system in the tree uniqueness region [16], and for colorings [15, 23] it matched the best known parameter bounds using other algorithmic approaches. Moreover, recent work of Chen et al. [17] shows that spectral independence implies optimal mixing of the Glauber dynamics in all of these cases as stated in the following theorem.

THEOREM 1.3. ([17]) *For an arbitrary spin system on a graph of maximum degree* Δ, *if the system is* η*-spectrally independent and* b*-marginally bounded, then there exists a constant* $C = C(b, \eta, \Delta) > 0$ *such that the mixing time of the Glauber dynamics for the spin system is at most* $Cn \log n$ *where* n *is the number of vertices, and the modified log-Sobolev constant of the Glauber dynamics is at least* $1/(Cn)$. *Moreover, the constant* C *satisfies* $C = \left(\frac{\Delta}{b}\right)^{O(1+\frac{\eta}{b})}$.

We note that we obtain an incremental improvement in the mixing time bound in this theorem, improving the exponent in the constant C from $O(1 + \eta/b^2)$ (see Theorem 1.9 in [17]) to $O(1 + \eta/b)$.

The key step in the proof of Theorem 1.3 is the implication

$$\text{Spectral Independence} \implies \text{Approximate Tensorization of Entropy.} \tag{1.2}$$

Approximate tensorization of entropy is a generalization of the standard fact that if μ is a product distribution, then for any function $f : \Omega \to \mathbb{R}_+$ we have $\mathrm{Ent}(f) \leq \sum_{x \in V} \mu[\mathrm{Ent}_x(f)]$, where $\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$ and $\mathrm{Ent}(f) = \mu[f \log(f/\mu[f])]$ denote the mean and entropy of f with respect to the measure μ, respectively, whereas $\mu[\mathrm{Ent}_x f] = \mu[f \log(f/\mu_x[f])]$ is the expected value according to μ of the conditional entropy $\tau \mapsto \mathrm{Ent}(f|\tau)$ for τ a spin configuration on $V \setminus \{x\}$; see Section 2.2 for the detailed definitions.

For a general Gibbs distribution μ, approximate tensorization of entropy says that there exists a constant $C \geq 1$, such that for any function $f : \Omega \to \mathbb{R}_+$,

$$\mathrm{Ent}(f) \leq C \sum_{x \in V} \mu[\mathrm{Ent}_x(f)]. \tag{1.3}$$

In general, approximate tensorization is easily seen to imply the desired bounds on the modified log-Sobolev constant and the mixing time of the Glauber dynamics; see e.g. [11]. In the setting of spin systems on the lattice $\mathbb{Z}^d$, approximate tensorization estimates are known to hold under the so-called strong spatial mixing condition; this follows from the logarithmic Sobolev inequalities established in [41, 34, 13].

We provide an alternative proof of some of the key steps for the implication (1.2). The analogous result in [17] is proved in the more general framework of simplicial complexes and generalizes the result of [18] for homogeneous strongly log-concave distributions; see also [29] for related results. Our proof, which is provided in the full version [4], is completely framed in the setting of spin systems and is devoid of any work on simplicial complexes. This new approach may be conceptually simpler to some readers, and it enables us to present a self-contained proof of our main results. As a byproduct we obtain the aforementioned improvement in the constant C in Theorem 1.3.

One of our main results in this paper is the following substantial extension of (1.2):

$$\text{Spectral Independence} \implies \text{General Block Factorization of Entropy}. \tag{1.4}$$

Caputo and Parisi [12] introduced the notion of *general block factorization* of entropy which generalizes approximate tensorization, and is useful for analyzing more general classes of Markov chains. Let $\alpha = (\alpha_B)_{B \subset V}$ be an arbitrary probability distribution over subsets of V, and set $\delta(\alpha) = \min_{x \in V} \sum_{B : B \ni x} \alpha_B$ as in (1.1). General block factorization of entropy holds with constant C if for all weights α, for all $f : \Omega \to \mathbb{R}_+$:

$$\delta(\alpha)\, \mathrm{Ent}\, f \leq C \sum_{B \subset V} \alpha_B\, \mu[\mathrm{Ent}_B\, f], \tag{1.5}$$

where $\mu[\mathrm{Ent}_B\, f] = \mu[f \log(f/\mu_B f)]$ is the expected value of the conditional entropy $\tau \mapsto \mathrm{Ent}(f|\tau)$ for τ a spin configuration on $V \setminus B$. Approximate entropy tensorization (1.3) is the special case when $\alpha_B = 1/n$ for every block of size 1 and $\alpha_B = 0$ for larger blocks. The choice of the constant $\delta(\alpha)$ in this inequality is motivated by the fact that when μ is a product measure then (1.5) holds with $C = 1$, in which case it is known as the Shearer inequality; see [11]. The block factorization of entropy is a statement concerning the equilibrium distribution μ which has deep consequences for several natural sampling algorithms. In particular, it implies optimal mixing and optimal entropy decay for arbitrary block dynamics and constitutes a key concept in the proof of Theorem 1.1 and Theorem 1.2. The precise formulation of (1.4) and its corollaries is as follows.

THEOREM 1.4. *For an arbitrary spin system on a graph of maximum degree Δ, if the system is η-spectrally independent and b-marginally bounded, then general block factorization of entropy (1.5) holds with constant $C = C(b, \eta, \Delta)$. Moreover, all heat-bath block dynamics have optimal mixing and optimal entropy decay. The constant C satisfies $C = \left(\frac{2}{b}\right)^{O\left(\Delta(1+\frac{\eta}{b})\right)}$.*

Recall, for the Glauber dynamics $\delta(\alpha) = 1/n$, and hence we recover Theorem 1.3 as a special case of the above result. As another example, for a bipartite graph, Theorem 1.4 implies $O(\log n)$ mixing time of the even-odd dynamics.

When the spin system satisfies (1.5) with α the uniform distribution over all subsets of a given size ℓ we refer to this as *ℓ-uniform block factorization of entropy* or ℓ-UBF for short. In [17], an important step in the proof of Theorem 1.3 is establishing ℓ-UBF with $\ell \sim \theta n$ for some $\theta \in (0, 1)$. To prove Theorem 1.4 for arbitrary blocks we establish that ℓ-UBF implies general block factorization of entropy; see Theorem 1.9 for a detailed statement and Figure 1 for a high-level overview.

Recent work of Blanca et al. [5] utilizes block factorization of entropy into the even and odd sublattices of $\mathbb{Z}^d$ to obtain tight mixing time bounds for the Swendsen-Wang dynamics on boxes of $\mathbb{Z}^d$ in the high-temperature region. Following the approach presented in [5] and using our general result in Theorem 1.4, here we prove optimal mixing time of the Swendsen-Wang dynamics when spectral independence holds on arbitrary bounded-degree graphs. This can be formalized in the following statement, which is a key ingredient in the proof of Theorem 1.2.

THEOREM 1.5. *For the ferromagnetic Ising and Potts models on a graph of maximum degree Δ, if the system is η-spectrally independent and b-marginally bounded, then there exists a constant $C = C(b, \eta, \Delta)$ such that the mixing time of the Swendsen-Wang dynamics is at most $C \log n$ and the modified log-Sobolev constant is at least C^{-1}. The constant C satisfies $C = \left(\frac{2}{b}\right)^{O\left(\Delta(1+\frac{\eta}{b})\right)}$.*

1.4 Establishing spectral independence The above results show the power of spectral independence as it implies optimal mixing time bounds for a wide variety of Markov chains. We next address when spectral independence holds and how it relates to classical conditions that imply fast mixing. The next series of results prove in a general context that when there exists a contractive coupling then spectral independence holds.

Let d denote an arbitrary metric on Ω. A simple example is the *Hamming metric*, which for configurations $\sigma, \tau \in \Omega$ is defined to be $d_{\mathrm{H}}(\sigma, \tau) = |\{x \in V : \sigma_x \neq \tau_x\}|$. There are two types of more general metrics that we will consider: those within a constant factor of the Hamming metric and vertex-weighted Hamming metric for arbitrary weights. For $\gamma \geq 1$, a metric d on Ω is said to be γ-equivalent to the Hamming metric (or γ-equivalent for simplicity) if for all $\sigma, \tau \in \Omega$,

$$\frac{1}{\gamma} d_{\mathrm{H}}(\sigma, \tau) \leq d(\sigma, \tau) \leq \gamma d_{\mathrm{H}}(\sigma, \tau);$$

that is, a γ-equivalent metric is an arbitrary metric where every distance is within a factor γ of the Hamming distance. In contrast, we can generalize the Hamming distance by considering arbitrary weights for the vertices. Let $w : V \to \mathbb{R}_+$ be an arbitrary positive weight function. The *w-weighted Hamming metric* between two configurations $\sigma, \tau \in \Omega$ is defined to be

$$d_w(\sigma, \tau) = \sum_{x \in V} w(x) \mathbf{1}\{\sigma_x \neq \tau_x\}.$$

In particular, if $w_x = 1$ for all x then d_w is just the usual Hamming metric. Note there are no constraints on the weights except that they are positive; in particular, the weights can be a function of n.

We will often consider a class $\mathcal{P} = \{P^\tau : \tau \in \mathcal{T}\}$ of Markov chains associated with μ, where each P^τ is a Markov chain with stationary distribution μ^τ and $\tau \in \mathcal{T}$ is a pinning; for example, $\mathcal{P}$ can be the family of Glauber dynamics for all μ^τ's. In coupling proofs, the goal is to design a coupling so that for an arbitrary pair of states the chains contract with respect to some distance metric after the coupled transition. Roughly speaking, for $\kappa \in (0, 1)$, we say that μ is κ-contractive with respect to (w.r.t.) a collection $\mathcal{P}$ of Markov chains and a metric d if one step of every chain P^τ contracts the distance by a factor κ in expectation. This is formalized in the following definition.

DEFINITION 1.4. (κ-CONTRACTION) *Let $\mathcal{P}$ denote a collection of Markov chains associated with μ and let d be a metric on Ω. For $\kappa \in (0, 1)$ we say that μ is κ-contractive w.r.t. $\mathcal{P}$ and d if for all $\tau \in \mathcal{T}$, all $X_0, Y_0 \in \Omega^\tau$, there exists a coupling $(X_0, Y_0) \to (X_1, Y_1)$ for P^τ such that:*

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

The following result shows that spectral independence holds if the Glauber dynamics has a contractive coupling.

THEOREM 1.6.

(1) If μ is κ-contractive w.r.t. the Glauber dynamics and an arbitrary w-weighted Hamming metric, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\epsilon}{n}$, then $\eta \leq \frac{2}{\epsilon}$.

(2) If the metric in (1) is not a weighted Hamming metric but instead an arbitrary γ-equivalent metric, then $\eta = \frac{2\gamma^2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\epsilon}{n}$, then $\eta \leq \frac{2\gamma^2}{\epsilon}$.

Note a κ-contractive coupling for the Hamming distance immediately implies $O(n \log n)$ mixing time of the Glauber dynamics (see, e.g., [10, 31]). But the above theorem offers two additional features. First, it allows arbitrary weights w and the resulting bound on the mixing time does not depend on the ratio of $\max_x w(x) / \min_x w(x)$, whereas a coupling argument, such as the one utilized in path coupling [10], yields a mixing

time bound which depends on this ratio. Second, as discussed in the previous theorems, spectral independence (together with the easily satisfied marginal boundedness) implies optimal bounds on the mixing time and entropy decay rate for arbitrary heat-bath block dynamics.

We can extend Theorem 1.6 by replacing the Glauber dynamics with arbitrary Markov chains. In particular, we consider a general class of Markov chains which we call the *select-update dynamics*. In each step, the select-update dynamics picks a block $B \in \mathcal{B}$ randomly (with a distribution that may depend on the current configuration), and updates all vertices in B using the current configuration (and the pinning if there is one). Note that no assumptions are made on how to pick or update the blocks; the only requirement is that the dynamics converges to the correct stationary distribution. If the chain selects a block B from a fixed distribution over $\mathcal{B}$ and updates B using the conditional marginal distribution on B (under the pinning if applicable), then this is the standard heat-bath block dynamics that we introduced earlier; hence, the select-update dynamics is much more general than the weighted heat-bath block dynamics. Another example of the select-update dynamics is the flip dynamics for sampling random colorings of a graph; see Section 3.3.1.

We define $M = \max_{B\in\mathcal{B}} |B|$ to be the maximum block size and D to be the maximum probability of a vertex being selected in any step of the chain; see (3.28) for the precise definition of D.

THEOREM 1.7. *If μ is κ-contractive w.r.t. arbitrary select-update dynamics and an arbitrary γ-equivalent metric, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2 DM}{1-\kappa}$.*

Theorem 1.7 generalizes Theorem 1.6(2) since $M = 1$ and $D = 1/n$ for the Glauber dynamics. If we further assume that the select-update dynamics updates each connected component of a block independently, then the maximum block size M can be replaced by the maximum component size of a block; see Remark 3.4. See also Theorem 3.4 for a stronger statement involving arbitrary Markov chains, where DM is replaced by the maximum expected distance of two chains when pinning a single vertex. This more general statement potentially applies to chains with unbounded block sizes, including the Swendsen-Wang dynamics.

It is worth remarking that, as a corollary of Theorem 1.7 we obtain that a coupling argument for the select-update dynamics where the maximum block size is constant (and $D/(1-\kappa) = O(1)$) implies $O(n \log n)$ mixing time of the Glauber dynamics, together with the optimal mixing and optimal entropy decay for arbitrary heat-bath block dynamics.

Moreover, as a corollary of Theorem 1.6 we obtain that the Dobrushin uniqueness condition implies spectral independence. The Dobrushin uniqueness condition is a classical condition in statistical physics which considers the following dependency matrix.

DEFINITION 1.5. (DOBRUSHIN UNIQUENESS CONDITION) *The* Dobrushin dependency (or influence) matrix $R \in \mathbb{R}^{V\times V}$ *is defined by* $R(x,x) = 0$ *and*

$$R(x,y) = \max\left\{d_{\mathrm{TV}}\left(\mu_y(\cdot \mid \sigma), \mu_y(\cdot \mid \tau)\right) : (\sigma,\tau) \in \mathcal{S}_{x,y}\right\} \qquad \text{for } x \neq y$$

where $\mathcal{S}_{x,y}$ is the set of all pairs of configurations on $V \setminus \{y\}$ that can differ only at x. The Dobrushin uniqueness condition *holds if the maximum column sum of R is at most $1-\epsilon$ for some $\epsilon > 0$.*

The Dobrushin dependency matrix for the entry $R(x,y)$ considers the worst case pair of configurations on the entire neighborhood of y which differ at x. If x is not a neighbor of y then $R(x,y) = 0$. Hence, the Dobrushin uniqueness condition states that for all y, $\sum_{x\in N(y)} R(x,y) < 1$. In contrast, the ALO influence matrix considers the influence of a disagreement at x on a vertex y (which is not necessarily a neighbor) and no other vertices are fixed, although one needs to consider all pinnings to establish spectral independence, so the notions are incomparable at first glance.

Using Theorem 1.6 we prove that the Dobrushin uniqueness condition implies spectral independence. Moreover, our result holds under generalizations of the Dobrushin uniqueness condition. Hayes [27] generalized it to the following spectral condition: if $\|R\|_2 \leq 1-\epsilon$ for some $\epsilon > 0$, then the mixing time of the Glauber dynamics is $O(n \log n)$. This was further generalized by Dyer et al. [20] to arbitrary matrix norms. We prove spectral independence when the spectral radius $\varrho(R) < 1$, which is the strongest statement of this type as the spectral radius is no larger than any matrix norm; see Remark 3.2 for a more detailed discussion.

THEOREM 1.8. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \leq 1-\epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

Previously, Marton [35] (see also [24, 40]) showed that the spectral condition in Theorem 1.8 implies approximate tensorization of entropy and thus optimal bounds on the modified log-Sobolev constant for the Glauber dynamics. See also [3] for related results with an alternative technique. However, the approach in these works does not imply block factorization of entropy as in our case.

REMARK 1.1. *Our definition of κ-contraction is equivalent to the statement that the Markov chain has coarse Ollivier-Ricci curvature at least $1-\kappa > 0$ with respect to the metric d [38]. Combining Theorem 1.6 with Theorem 1.4 we obtain a proof of the following version of the Peres-Tetali conjecture: if the Glauber dynamics has Ollivier-Ricci curvature at least $\epsilon/n > 0$ then the Glauber dynamics has a modified log-Sobolev constant at least c/n and any α-weighted heat-bath block dynamics has a modified log-Sobolev constant at least $c\,\delta(\alpha)$, for some constant $c = c(\epsilon, b, \Delta) > 0$, where $\delta(\alpha)$ is defined in (1.1). Replacing Theorem 1.6 with its generalization Theorem 1.7 we obtain the same conclusion under the much milder assumption that there exists some κ-contractive select-update dynamics satisfying $DM/(1-\kappa) = O(1)$. The original Peres-Tetali conjecture in the setting of random walks on graphs is that if there exists a graph metric d such that the random walk has Ollivier-Ricci curvature at least $\lambda > 0$ with respect to d then the random walk has modified log-Sobolev constant at least $c\lambda > 0$, for some universal constant $c > 0$; see Conjecture 3.1 in Eldan et al. [22].*

1.5 Proof overviews for main results In this section, we sketch the proofs of our key technical results. We begin in Section 1.5.1 with an overview of our proofs that spectral independence implies optimal bounds for arbitrary block dynamics and the Swendsen-Wang dynamics (namely, Theorems 1.4 and 1.5). In Section 1.5.2 we highlight the proofs of Theorems 1.6, and 1.7 that a contractive coupling for an arbitrary local dynamics implies spectral independence.

1.5.1 Optimal mixing under spectral independence: Theorems 1.4 and 1.5 We begin with the high-level idea for the proof that spectral independence implies optimal mixing for arbitrary heat-bath block dynamics, and then we describe the key ideas to obtain optimal mixing for the Swendsen-Wang dynamics.

Recall from Section 1.3 that to establish optimal mixing for an arbitrary choice of block dynamics it suffices to prove general block factorization (GBF); see (1.5). Previous results concerning spectral independence show that spectral independence implies ℓ-uniform block factorization (ℓ-UBF) with $\ell = \lceil \theta n \rceil$ and $\theta \in (0,1)$. Note, ℓ-UBF refers to the block factorization where the weights α are uniform over all subsets of size ℓ; see Definition 1.6 below. The key step in the proof of Theorem 1.4 is to show that ℓ-UBF, with $\ell = \lceil \theta n \rceil$ and θ sufficiently small, implies general block factorization (GBF).

We begin with the formal definition of ℓ-UBF. For a positive integer $\ell \le n$, let $\binom{V}{\ell}$ denote the collection of all subsets of V of size ℓ.

DEFINITION 1.6. (UNIFORM BLOCK FACTORIZATION (UBF)) *We say that the spin system μ satisfies the ℓ-uniform block factorization (ℓ-UBF) of entropy with constant C_{UBF} if for all $f : \Omega \to \mathbb{R}_+$*

$$\frac{\ell}{n}\,\mathrm{Ent}(f) \le C_{\text{UBF}} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\mathrm{Ent}_S(f)]. \tag{1.6}$$

We prove the following theorem that ℓ-UBF (for sufficiently small choice of ℓ) implies general block factorization (GBF).

THEOREM 1.9. *For an arbitrary b-marginally bounded spin system on a graph of maximum degree Δ, if $\lceil \theta n \rceil$-UBF holds with constant C_{UBF} and $0 < \theta \le \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, then GBF holds with constant $C_{\text{GBF}} = C_{\text{UBF}} \times O\left((\theta\, b^2)^{-1} \log(1/b)\Delta^3\right)$.*

As mentioned earlier, spectral independence is known to imply $\lceil \theta n \rceil$-UBF for any desired constant $\theta > 0$ [17]. These two results combined imply Theorem 1.4 from the introduction:

Proof of Theorem 1.4. For a spin system that is η-spectrally independent and b-marginally bounded, $\lceil \theta n \rceil$-UBF holds with constant $C_{\text{UBF}} = (\frac{1}{\theta})^{O(\frac{\eta}{b})}$ (see Theorem 5.1 in [4]). Then, taking $\theta = \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, Theorem 1.9 implies that GBF holds with constant $C_{\text{GBF}} = O\left(\frac{4e\Delta^5}{b^{2(\Delta+2)}} \log(1/b)\right) \times \left(\frac{4e\Delta^2}{b^{2(\Delta+1)}}\right)^{O\left(\frac{\eta}{b}\right)}$, and it thus follows that

$C_{\text{GBF}} = \left(\frac{2}{b}\right)^{O\left(\Delta\left(1+\frac{\eta}{b}\right)\right)}$. □

Hence, the key novelty in the proof of Theorem 1.4 is Theorem 1.9. To establish Theorem 1.9 we consider a special case of GBF, which we call *k-partite factorization of entropy*. Recall that a graph G of maximum degree Δ is k-partite, with $k \le \Delta+1$. Let $\{V_1, ..., V_k\}$ denote the independent sets $V_i \subset V$ corresponding to a k-partition of G. Theorem 1.9 follows immediately from the following factorization statements.

LEMMA 1.1. *Suppose that for an arbitrary b-marginally bounded spin system on a graph of maximum degree Δ, $\lceil \theta n \rceil$-UBF holds with constant C_{UBF} and $\theta \le \frac{b^{2(\Delta+1)}}{4e\Delta^2}$. Then,*

$$\text{Ent}(f) \le K C_{\text{UBF}} \sum_{i=1}^{k} \mu[\text{Ent}_{V_i}(f)], \tag{1.7}$$

where the constant K satisfies $K = O(\Delta^2(\theta\, b^2)^{-1} \log(1/b))$. We refer to inequality (1.7) as a k-partite factorization of entropy with constant KC_{UBF}.

LEMMA 1.2. *Suppose that for an arbitrary spin system on a graph of maximum degree Δ, k-partite factorization of entropy holds with constant C. Then, GBF holds with constant Ck.*

We comment briefly on how we prove these two lemmas; their actual proofs can be found in the full version of the paper [4]. The main idea behind the proof of Lemma 1.1 can be roughly explained as follows. If the sets S in (1.6) were all independent sets, then standard results about entropy factorization under product distributions would yield the desired conclusion. Moreover, the same conclusion would continue to hold if S only contained connected components of constant size. However, even if θ is sufficiently small, the random set S is likely to have components of size $\Theta(\log n)$, but the expected component size is $O(1)$. The challenge in obtaining optimal bounds is to utilize the expected instead of the maximum component size. The proof of Lemma 1.2, on the other hand, relies on the fact that GBF holds on each of the independent sets V_i; this is a consequence of the weighted Shearer inequality for the Shannon entropy (see, e.g., Lemma 4.2 in [12]).

Finally, to prove Theorem 1.5 (i.e., the optimal mixing results for the SW dynamics), a key concept in our strategy is the *spin/edge factorization* of entropy in the "joint" Edwards-Sokal space [21, 25], the set of spin-edge configurations consisting of a spin assignment to the vertices and a subset of edges compatible with the spin assignment (i.e., every edge is monochromatic). (The SW dynamics alternates between spin and spin-edge configurations.) The spin/edge factorization of entropy was shown in [5, Lemma 1.8] to imply $O(\log n)$ mixing of the SW dynamics on any graph. To prove Theorem 1.5, we show that k-partite factorization of entropy for μ implies spin/edge factorization of entropy. This implication was established in [5] for bipartite graphs, and we extend it to the case of general graphs. The proof of Theorem 1.5 is provided in the full version [4].

1.5.2 Spectral independence via contractivity: Theorems 1.6 and 1.7 Here we outline our proofs of Theorems 1.6 and 1.7. We establish spectral independence by showing that the maximum absolute row sum of the ALO influence matrix is bounded, as done in previous works [2, 17, 15, 23]. Consider the case without pinnings for simplicity. We would like to upper bound, for each $(x, a) \in \mathcal{X}$, the quantity

$$S(x,a) = \sum_{(y,a')\in\mathcal{X}} |J(x,a;y,a')| = \sum_{(y,a')\in\mathcal{X}:\, y\neq x} |\nu(\sigma_y = a') - \mu(\sigma_y = a')|$$

where $\nu = \mu(\cdot \mid \sigma_x = a)$ is the conditional distribution under the pinning $\sigma_x = a$. Upper bounds on $S(x,a)$ (and analogous results with pinnings) would then imply spectral independence. With some effort, we can define a 2-Lipschitz function $f : \Omega \to \mathbb{R}$ (w.r.t. the Hamming metric d_{H}) such that $S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f$. In particular, we have $S(x,a) \le 2W_{1,d_{\text{H}}}(\nu,\mu)$ where $W_{1,d_{\text{H}}}(\nu,\mu)$ represents the 1-Wasserstein distance w.r.t. the Hamming metric; we refer to Section 3.1 for relevant definitions. The important intuition here is that it suffices to upper bound some statistical distance between the two distributions μ and $\nu = \mu(\cdot \mid \sigma_x = a)$. In other words, to deduce spectral independence one only needs to show that every pinning $\sigma_x = a$ would disturb the distribution μ on a limited scale, in terms of the Wasserstein distance.

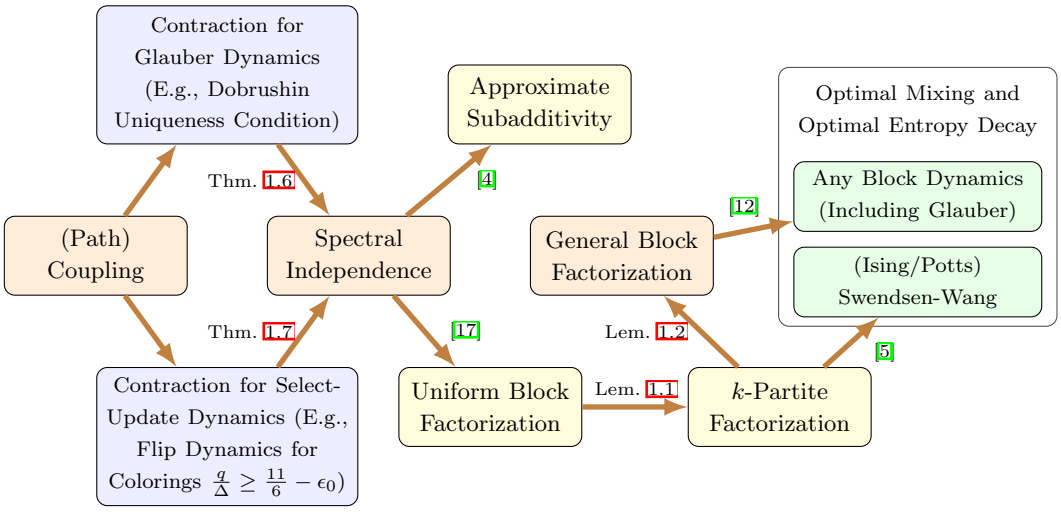

Figure 1: Organization of the paper

Up to now, we have not yet applied our assumptions on contractivity of the distribution μ. Our next step is to show $W_{1,d_{\mathrm{H}}}(\nu,\mu) = O(1)$ for contractive μ. To achieve this, we generalize a result from previous works [9, 39] to bound the Wasserstein distance of two distributions; see Lemma 3.1. Roughly speaking, we show that, assuming contractivity, the stationary distributions of two Markov chains are close to each other if the two chains are close in one step. Previous results in [9, 39] were specialized for the binary product space and the Glauber dynamics. Here, we establish our Lemma 3.1 for any finite state space and any Markov chain. This result is of independent interest and may find applications in other problems.

Independent work. Independently, [32] established similar results to Theorem 1.7 which also implied Theorem 1.1 for the Glauber dynamics. In [32] the author also concluded a version of Theorem 1.8 but required the stronger assumption that the row sum of the Dobrushin dependency matrix is bounded. In contrast, using our Theorem 1.6 we only require a bound on the spectral radius which is the weakest assumption of this type; see Remark 3.2.

Paper organization. The organization of the paper is demonstrated in Figure 1. We provide the necessary background in Section 2. In Section 3, we then establish spectral independence if the distribution admits a contractive Markov chain. All other proofs are deferred to the full version of the paper [4].

2 Preliminaries

2.1 Spin systems We begin with the formal definition of general q-state spin systems. Let $q \geq 2$ be an integer and write $[q] = \{1, \dots, q\}$. Let $G = (V \cup \partial V, E \cup \partial E)$ be an undirected graph where ∂V denotes the boundary set of the induced subgraph $G' = (V, E)$, and ∂E consists of all edges between V and ∂V. A q-spin system on G with a boundary condition $\xi \in [q]^{\partial V}$ is parameterized by nonnegative symmetric matrices $A_{xy} \in \mathbb{R}_+^{q\times q}$, $\{x, y\} \in E \cup \partial E$, representing the nearest neighbor interactions, and vectors $B_x \in \mathbb{R}_+^q$, $x \in V$, representing the external fields. A configuration $\sigma \in [q]^V$ has weight:

$$w(\sigma) = \prod_{\{x,y\}\in E} A_{xy}(\sigma_x, \sigma_y) \prod_{\substack{\{x,y\}\in\partial E \\ x\in V,\, y\in\partial V}} A_{xy}(\sigma_x, \xi_y) \prod_{x\in V} B_x(\sigma_x).$$

Let $\Omega = \{\sigma \in [q]^V : w(\sigma) > 0\}$ denote the collection of all feasible configurations and let $Z_G = \sum_{\sigma\in\Omega} w(\sigma)$ denote the partition function. We assume that $\Omega \neq \emptyset$; i.e., the boundary condition ξ is feasible. Finally, the *Gibbs distribution* μ is given by, for $\sigma \in \Omega$,

$$\mu(\sigma) = w(\sigma)/Z_G.$$

Recall the notion of pinning from Section 1.2 which we briefly repeat here for convenience. For $U \subset V$, we use the notation $\sigma_U = (\sigma_x)_{x\in U}$ and let $\Omega_U = \{\tau \in [q]^U : \exists \sigma \in \Omega, \sigma_U = \tau\}$ be the set of all possible pinnings on U. Note, for $x \in V$, Ω_x is the set of feasible spin assignments for vertex x. Denote the collection of all pinnings by $\mathcal{T} = \cup_{U\subset V}\Omega_U$ and denote the set of all feasible vertex-spin pairs by $\mathcal{X} = \{(x, a) : x \in V, a \in \Omega_x\}$. For $\tau \in \mathcal{T}$, let μ^τ denote the conditional Gibbs distribution $\mu(\cdot \,|\, \sigma_U = \tau)$. We also write $\mu^\tau_\Lambda = \mu^\tau$ if $\tau \in \Omega_{V\setminus\Lambda}$ and use the notation $\mu_\Lambda : \Omega_{V\setminus\Lambda} \ni \tau \mapsto \mu^\tau_\Lambda$ for the associated mapping.

For a pinning $\tau \in \Omega_U$ for $U \subset V$, let $\Omega^\tau = \{\sigma \in \Omega : \sigma_U = \tau\}$ denote the corresponding state space; i.e., Ω^τ is the support of μ^τ. We also define $\Omega^\tau_W = \{\varphi \in [q]^W : \exists \sigma \in \Omega^\tau, \sigma_W = \varphi\}$ for $W \subset V \setminus U$ and $\mathcal{X}^\tau = \{(x, a) : x \in V \setminus U, a \in \Omega^\tau_x\}$. We say Ω^τ is connected if the graph on Ω^τ with edges connecting pairs at Hamming distance 1 is connected. The distribution μ over Ω is said to be *totally-connected* if for every $\tau \in \mathcal{T}$, the set Ω^τ is connected. Throughout this paper, we will assume the distribution μ is totally-connected as this is necessary for the Glauber dynamics to be ergodic for all conditional measures μ^τ.

We recall some classical examples of spin system. The Ising/Potts model at inverse temperature $\beta \in \mathbb{R}$ corresponds to the interaction $A_{xy}(a, a') = \exp(\beta \mathbf{1}(a = a'))$ and $B_x(a) = \exp(h(a))$ where $h \in \mathbb{R}^q$ is a vector of external fields, with $q = 2$ for the Ising model and $q \geq 3$ for the Potts model. The hard-core (or independent sets) model with parameter $\lambda > 0$ is obtained with $q = 2$, $A_{xy}(a, a') = 0$ if $a = a' = 1$ and $A_{xy}(a, a') = 1$ otherwise, and $B_x(a) = \lambda$ if $a = 1$ and $B_x(a) = 1$ if $a = 2$. The q-colorings model corresponds to $A_{xy}(a, a') = \mathbf{1}(a \neq a')$ and $B_x(a) = 1$. Note that the Ising/Potts models with any β and h, as well as the hard-core model with any $\lambda > 0$, and the q-colorings when $q \geq \Delta + 2$ are totally-connected spin systems.

2.2 Mixing time, entropy, and log-Sobolev inequalities Let P be the transition matrix of an ergodic Markov chain with finite state space Ω and stationary distribution μ. Let $P^t(X_0, \cdot)$ denote the distribution of the chain after t steps starting from the initial state $X_0 \in \Omega$. The *mixing time* $T_{\text{mix}}(P)$ of the chain is defined as

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

where $\|\cdot\|_{\text{TV}}$ denotes total variation distance.

In this paper, we rely on functional inequalities related to entropy to bound the mixing time. For a function $f : \Omega \mapsto \mathbb{R}$, let $\mu[f] = \sum_{\sigma\in\Omega} \mu(\sigma)f(\sigma)$ and $\text{Var}_\mu(f) = \mu[f^2] - \mu[f]^2$ denote its mean and variance with respect to μ. Likewise, for $f : \Omega \to \mathbb{R}_+$, the *entropy* of f with respect to μ is defined as

$$\text{Ent}(f) = \mu\left[f \cdot \log\left(\frac{f}{\mu[f]}\right)\right] = \mu[f \cdot \log f] - \mu[f] \cdot \log \mu[f]. \tag{2.8}$$

When $f \geq 0$ is such that $\mu[f] = 1$, then $\text{Ent}(f) = H(f\mu \,|\, \mu)$ equals the *relative entropy*, or Kullback-Leibler divergence, of the distribution $f\mu$ with respect to μ.

For real functions f, g on Ω, the *Dirichlet form* associated to the pair (P, μ) is defined as

$$\mathcal{D}_P(f, g) = \langle f, (1 - P)g\rangle_\mu, \tag{2.9}$$

where $\langle f, g\rangle_\mu = \mu[fg]$ denotes the scalar product in $L^2(\mu)$. When P is reversible, i.e., $\mu(\sigma)P(\sigma, \tau) = \mu(\tau)P(\tau, \sigma)$, one has

$$\mathcal{D}_P(f, g) = \frac{1}{2}\sum_{\sigma,\tau\in\Omega} \mu(\sigma)P(\sigma, \tau)(f(\sigma) - f(\tau))(g(\sigma) - g(\tau)). \tag{2.10}$$

Definition 2.1. *The pair (P, μ) satisfies the (standard)* log-Sobolev inequality *(LSI) with constant s if for all $f \geq 0$:*

$$\mathcal{D}_P(\sqrt{f}, \sqrt{f}) \geq s\,\text{Ent}(f). \tag{2.11}$$

It satisfies the modified log-Sobolev inequality *(MLSI) with constant ϱ if for all $f \geq 0$:*

$$\mathcal{D}_P(f, \log f) \geq \varrho\,\text{Ent}(f). \tag{2.12}$$

It satisfies the (discrete time) relative entropy decay with rate $\delta > 0$ *if for all distributions ν:*

$$H(\nu P \,|\, \mu) \leq (1 - \delta)H(\nu \,|\, \mu). \tag{2.13}$$

In this paper we focus on the entropy decay inequality (2.13) which may be seen as the discrete time analog of the modified log-Sobolev inequality. We recall some well known facts about its relation to the other two inequalities and its implications for mixing times.

LEMMA 2.1. *If (P,μ) satisfies the standard LSI with constant s then it satisfies the MLSI with constant $\varrho = 2s$. If it satisfies the discrete time relative entropy decay with rate $\delta > 0$, then it satisfies the MLSI with constant $\varrho = \delta$. Finally, if it satisfies the discrete time relative entropy decay with rate $\delta > 0$, then*

$$T_{\mathrm{mix}}(P) \le 1 + \delta^{-1}[\log(8) + \log\log(1/\mu_*)]\,, \tag{2.14}$$

where $\mu_* = \min_{\sigma\in\Omega}\mu(\sigma)$.

We refer to e.g. [5, Section 2] for a proof. Note that we have not assumed reversibility of P in the above lemma. If (P,μ) is reversible, then one can additionally show that the standard LSI with constant s implies the discrete time relative relative entropy decay with rate $\delta = s$.

2.3 Implications of block factorization Fix a probability distribution α over subsets of V and observe that the α-weighted heat bath block dynamics defined in the introduction is the Markov chain with transition matrix P_α on Ω such that for any real function f

$$P_\alpha f = \sum_{B\subset V} \alpha_B\, \mu_B(f)\,. \tag{2.15}$$

To clarify the above notation, if we evaluate the left hand side at a spin configuration $\sigma\in\Omega$ then each for each B the term $\mu_B f$ in the right hand side is given by $\mu_B^\tau f$ where $\tau = \sigma_{V\setminus B}$. If $\alpha_B = n^{-1}\,\mathbf{1}(|B|=1)$, then (2.15) is the Glauber dynamics for μ.

The α-weighted heat bath block dynamics (2.15) defines a reversible pair (P_α,μ). Moreover, its Dirichlet form satisfies

$$\mathcal{D}_\alpha(f,g) = \sum_{B\subset V}\alpha_B\,\mu[f(1-\mu_B)g] = \sum_{B\subset V}\alpha_B\,\mu\left[\mathrm{Cov}_B(f,g)\right]\,, \tag{2.16}$$

where $\mathrm{Cov}_B(f,g) = \mu_B\left[(f-\mu_B f)(g-\mu_B g)\right]$ denotes the covariance functional.

LEMMA 2.2. *If the spin system satisfies the general block factorization with constant C then for all α the Markov chain (P_α,μ) satisfies*

1. *the modified log-Sobolev inequality with constant $\varrho = \frac{\delta(\alpha)}{C}$;*
2. *the discrete time relative entropy decay with rate $\delta = \frac{\delta(\alpha)}{C}$;*
3. *$T_{\mathrm{mix}}(P_\alpha) \le 1 + \frac{C}{\delta(\alpha)}[\log(8) + \log\log(1/\mu_*)]$, where $\mu_* = \min_{\sigma\in\Omega}\mu(\sigma)$.*

A proof of this lemma is provided in the full version of the paper [4].

3 Spectral independence for contractive distributions

In this section we establish our main results that a contractive distribution is spectrally independent. These results in particular connect classic probabilistic approach for establishing fast mixing of Markov chains such as coupling with recent developments utilizing spectral independence. We first consider a special case of Theorem 1.6 concerned with Glauber dynamics and Hamming metric in Section 3.1; this will serve as a concrete example to illustrate our approach for establishing spectral independence. In Section 3.2, we consider arbitrary metric and prove Theorem 1.6. Finally, we consider general Markov chains and metrics in Section 3.3 and prove Theorem 1.7.

3.1 Warm-up: contraction for Glauber dynamics and Hamming metric In this section, we prove a simpler version of Theorem 1.6, which already gives the main idea of our proof approach for establishing spectral independence. We show that, if the distribution μ is contractive w.r.t. the Glauber dynamics and the Hamming metric, then it is spectrally independent.

THEOREM 3.1. *If μ is κ-contractive w.r.t. the Glauber dynamics and the Hamming metric for some $\kappa \in (0,1)$, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \le 1 - \epsilon/n$, then $\eta \le 2/\epsilon$.*

REMARK 3.1. *In this paper, the Glauber dynamics P^τ_{GL} for the conditional distribution μ^τ with a pinning τ on $U \subset V$ is defined as follows: in each step the chain picks a vertex $x \in V$ u.a.r. and updates its spin conditioned on all other vertices and τ. In particular, all pinned vertices in U are allowed to be selected and when this happens the configuration will remain the same (no updates will be made). This setting can make our theorem statements and proofs easier to understand, and will not harm our results since we only consider these chains for the purpose of analysis rather than actually running them. Alternatively, we can define the Glauber dynamics $\tilde{P}^\tau_{\mathrm{GL}}$ for μ^τ in the following way: in each step an* unpinned *vertex $x \in V \setminus U$ is selected u.a.r. and updated accordingly. Note that $\tilde{P}^\tau_{\mathrm{GL}}$ is faster than P^τ_{GL} and the contraction rate of $\tilde{P}^\tau_{\mathrm{GL}}$ depends on the number of unpinned vertices. If we assume μ^τ is κ_ℓ-contractive w.r.t. $\tilde{P}^\tau_{\mathrm{GL}}$ and d_{H} where $\ell = |V \setminus U|$, then an analog of Theorem 3.1 can show that μ is spectrally independent with*

$$\eta = \max_{\ell=1,\dots,n} \left\{ \frac{2}{(1-\kappa_\ell)\ell} \right\}.$$

However, in actual applications such as under the Dobrushin uniqueness condition in Section 3.2.1, the contraction rate satisfies $\kappa_\ell \le 1 - \epsilon/\ell$, so we eventually get $\eta \le 2/\epsilon$ just as from Theorem 3.1.

Recall that for any pinning $\tau \in \mathcal{T}$ we let μ^τ be the conditional distribution over Ω^τ given τ, and the ALO influence matrix J^τ is a square matrix indexed by $\mathcal{X}^\tau$ and defined as $J(x,a;x,a') = 0$ and

$$J^\tau(x,a;y,a') = \mu^\tau(\sigma_y = a' \mid \sigma_x = a) - \mu^\tau(\sigma_y = a') \text{ for } x \neq y.$$

The distribution μ is said to be η-spectrally independent if $\lambda_1(J^\tau) \le \eta$ for all pinning τ.

Our goal is to upper bound the maximum eigenvalue of the ALO influence matrix J^τ for a given pinning τ. In fact, to make notations simpler we will only consider the case where there is no pinning; the proof is identical by replacing Ω, μ, J with $\Omega^\tau, \mu^\tau, J^\tau$ when an arbitrary pinning τ is given. To upper bound $\lambda_1(J)$, a standard approach that has been applied in previous works [2, 16, 15, 23, 17] is to upper bound the infinity norm of J. More specifically, for each $(x,a) \in \mathcal{X}$ we define

$$S(x,a) = \sum_{(y,a') \in \mathcal{X}} |J(x,a;y,a')| \tag{3.17}$$

to be the sum of absolute influences of a given pair (x,a). The quantity $S(x,a)$ can be thought of as the total influence of (x,a) on all other vertex-spin pairs. If one can show $S(x,a) \le \eta$ for all $(x,a) \in \mathcal{X}$, then it immediately follows that

$$\lambda_1(J) \le \|J\|_\infty = \max_{(x,a)\in\mathcal{X}} S(x,a) \le \eta.$$

Hence, it suffices to show that $S(x,a) = O(1)$.

Fix $(x,a) \in \mathcal{X}$, and define the distribution $\nu = \mu(\cdot \mid \sigma_x = a)$; namely, ν is the conditional distribution of μ with the pinning $\sigma_x = a$. The key observation we make here is that the quantity $S(x,a)$ can be viewed as the difference of the expectation of some function f under the two measures μ and ν. More specifically, we define

$$f(\sigma) = \sum_{(y,a') \in \mathcal{X}} t(x,a;y,a')\, \mathbf{1}_{\{\sigma_y = a'\}}, \tag{3.18}$$

where

$$t(x,a;y,a') = \operatorname{sgn}(J(x,a;y,a')) = \begin{cases} +1, & J(x,a;y,a') > 0; \\ -1, & J(x,a;y,a') < 0; \\ 0, & J(x,a;y,a') = 0. \end{cases}$$

With this definition it follows that

$$\begin{aligned} S(x,a) &= \sum_{(y,a')\in\mathcal{X}} t(x,a;y,a')J(x,a;y,a') \\ &= \sum_{(y,a')\in\mathcal{X}} t(x,a;y,a')\mu(\sigma_y = a' \mid \sigma_x = a) - t(x,a;y,a')\mu(\sigma_y = a') \\ &= \mathbb{E}_\nu f - \mathbb{E}_\mu f. \end{aligned}$$

Therefore, the absolute sum of influences $S(x,a)$ describes, in some sense, the "distance" of the two distributions ν and μ measured by f.

To be more precise about our last statement, we review some standard definitions about the Wasserstein distance. Let (Ω, d) be a finite metric space. We say a function $f : \Omega \to \mathbb{R}$ is *L-Lipschitz* w.r.t. the metric d if for all $\sigma, \tau \in \Omega$ we have

$$|f(\sigma) - f(\tau)| \le L d(\sigma, \tau).$$

For every function $f : \Omega \to \mathbb{R}$, we let $L_d(f)$ be the optimal Lipschitz constant of f w.r.t. the metric d; i.e., $L_d(f) = \inf\{L \ge 0 : f \text{ is } L\text{-Lipschitz w.r.t. } d\}$. For a pair of distributions μ and ν on Ω, the 1-*Wasserstein distance* w.r.t. the metric d between μ and ν is defined as

$$W_{1,d}(\mu,\nu) = \inf_{\pi\in\mathcal{C}(\mu,\nu)} \mathbb{E}_\pi[d(\sigma,\tau)],$$

where $\mathcal{C}(\mu,\nu)$ denotes the set of all couplings of μ,ν (i.e., $\pi(\cdot,\cdot) \in \mathcal{C}(\mu,\nu)$ is a joint distribution over $\Omega\times\Omega$ with the marginals on the first and second coordinates being μ and ν respectively) and (σ,τ) is distributed as π; equivalently, the 1-Wasserstein distance can be represented as

$$W_{1,d}(\mu,\nu) = \sup_{\substack{f:\Omega\to\mathbb{R} \\ L_d(f)\le 1}} \mathbb{E}_\mu f - \mathbb{E}_\nu f. \tag{3.19}$$

Observe that, the function f defined by (3.18) is 2-Lipschitz w.r.t. the Hamming metric d_{H}; to see this, if $\sigma,\tau\in\Omega$ and $d_{\mathrm{H}}(\sigma,\tau) = k$ then by the definition of f we have $|f(\sigma) - f(\tau)| \le 2k$. Therefore, we deduce from (3.19) that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le L_{d_{\mathrm{H}}}(f)\, W_{1,d_{\mathrm{H}}}(\nu,\mu) \le 2W_{1,d_{\mathrm{H}}}(\nu,\mu).$$

That means, if one can show $W_{1,d_{\mathrm{H}}}(\nu,\mu) = O(1)$ for μ and $\nu = \mu(\cdot \mid \sigma_x = a)$ for any pair (x,a), then $\lambda_1(J) = O(1)$ and spectral independence would follow.

The following lemma, which generalizes previous works [9, 39], will be used to bound the Wasserstein distance of two distributions and may be interesting of its own. Roughly speaking, it claims that if μ,ν are the stationary distributions of two Markov chains P, Q (e.g., Glauber dynamics) respectively, and if μ is contractive w.r.t. P and the two chains P, Q are "close" to each other in one step, then the Wasserstein distance between ν and μ is small. The special case where $\Omega = \{+,-\}^n$ and P, Q are both the Glauber dynamics appeared in [9, Theorem 3.1] and [39, Theorem 2.1], but here we do not make any assumption on the state space or the chains, which is crucial to our applications in Section 3.3.1.

LEMMA 3.1. *Let (Ω, d) be a finite metric space. Let μ,ν be two distributions over Ω, and P, Q be two Markov chains on Ω with stationary distributions μ,ν respectively. If μ is κ-contractive w.r.t. the chain P and the metric d, then for every $f : \Omega \to \mathbb{R}$ we have*

$$|\mathbb{E}_\mu f - \mathbb{E}_\nu f| \le \frac{L_d(f)}{1-\kappa}\, \mathbb{E}_\nu\left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot))\right]$$

where $P(\sigma,\cdot)$ is the distribution after one step of the chain P when starting from σ and similarly for $Q(\sigma,\cdot)$. As a consequence,

$$W_{1,d}(\mu,\nu) \le \frac{1}{1-\kappa}\, \mathbb{E}_\nu\left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot))\right].$$

We remark that Lemma 3.1 holds in a very general setting, and (Ω, d) can be any finite metric space. It shows that if two Markov chains are close to each other, then their stationary distributions must be close to each other, under the assumption that one of the chains is contractive.

Proof of Lemma 3.1. The proof imitates the arguments from [9, 39]. Assume for now that P is irreducible; this is a conceptually easier case and we will consider general P later. Since P is irreducible, let h be the principal solution to the Poisson equation $(I-P)h = \bar{f}$ where $\bar{f} = f - \mathbb{E}_\mu f$; that is,

$$h = \sum_{t=0}^{\infty} P^t \bar{f}. \tag{3.20}$$

See Lemma 2.1 in [9] and the references in that paper for backgrounds on the Poisson equation. We then have

$$\mathbb{E}_\nu f - \mathbb{E}_\mu f = \mathbb{E}_\nu \bar{f} = \mathbb{E}_\nu[(I-P)h] = \mathbb{E}_\nu[(Q-P)h]$$

where the last equality is due to $\nu = \nu Q$. For each $\sigma \in \mathrm{supp}(\nu) \subset \Omega$, we deduce from (3.19) that

$$((Q-P)h)(\sigma) = \mathbb{E}_{Q(\sigma,\cdot)} h - \mathbb{E}_{P(\sigma,\cdot)} h \le L_d(h)\, W_{1,d}(Q(\sigma,\cdot), P(\sigma,\cdot)).$$

It remains to bound the Lipschitz constant of h. For $\sigma, \tau \in \Omega$,

$$\begin{aligned} |h(\sigma) - h(\tau)| &\le \sum_{t=0}^{\infty} \left| (P^t \bar{f})(\sigma) - (P^t \bar{f})(\tau) \right| \\ &= \sum_{t=0}^{\infty} \left| \mathbb{E}_{P^t(\sigma,\cdot)} \bar{f} - \mathbb{E}_{P^t(\tau,\cdot)} \bar{f} \right| \\ &\le L_d(f) \sum_{t=0}^{\infty} W_{1,d}(P^t(\sigma,\cdot), P^t(\tau,\cdot)) \end{aligned}$$

where the last inequality again follows from (3.19). Since μ is κ-contractive w.r.t. P and d, for all $\sigma, \tau \in \Omega$ and every integer $t \ge 1$ we have

$$W_{1,d}(P^t(\sigma,\cdot), P^t(\tau,\cdot)) \le \kappa^t d(\sigma,\tau).$$

We then deduce that

$$|h(\sigma) - h(\tau)| \le L_d(f) \sum_{t=0}^{\infty} \kappa^t d(\sigma,\tau) = \frac{L_d(f)}{1-\kappa}\, d(\sigma,\tau).$$

This implies that $L_d(h) \le L_d(f)/(1-\kappa)$ and the lemma then follows.

Next, we show how to remove the assumption that P is irreducible. Observe that in the proof above we only need the irreducibility of P to guarantee that the function h given by (3.20) is well-defined; i.e., the series on the right-hand side of (3.20) is convergent. The rest of the proof does not require the irreducibility of P. In fact, one can deduce the convergence of (3.20) solely from the contraction of P. Note that for all $\sigma \in \Omega$,

$$\begin{aligned} \left| P^t \bar{f}(\sigma) \right| &= \left| P^t \bar{f}(\sigma) - \mathbb{E}_\mu P^t \bar{f} \right| \\ &= \left| P^t \bar{f}(\sigma) - \sum_{\tau \in \Omega} \mu(\tau) P^t \bar{f}(\tau) \right| \\ &\le \sum_{\tau \in \Omega} \mu(\tau) \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right| \end{aligned}$$

where the first equality follows from $\mathbb{E}_\mu P^t \bar{f} = \mathbb{E}_\mu \bar{f} = 0$. Since Ω is finite, to show that (3.20) is convergent for all $\sigma \in \Omega$, it suffices to show that for all $\sigma, \tau \in \Omega$ the series $\sum_{t=0}^{\infty} \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right|$ is convergent. Actually, our proof before has already showed that

$$\sum_{t=0}^{\infty} \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right| \le \frac{L_d(f)}{1-\kappa}\, d(\sigma,\tau) < \infty$$

using only the contraction of P, where we have $L_d(f) < \infty$ and $\sup_{\sigma,\tau\in\Omega} d(\sigma,\tau) < \infty$ because Ω is finite. Therefore, the lemma remains true without the assumption of irreducibility of P. □

Given Lemma 3.1, we can now complete the proof of Theorem 3.1.

Proof of Theorem 3.1. For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 3.1 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le \frac{L_{d_{\mathrm{H}}}(f)}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot), Q(\sigma,\cdot))\right] \tag{3.21}$$

where $S(x,a)$ is given by (3.17), f is given by (3.18), P is the Glauber dynamics for μ, and Q is the Glauber dynamics for $\nu = \mu^{(x,a)} = \mu(\cdot \mid \sigma_x = a)$ (we use (x,a) to denote the pinning $\sigma_x = a$). We claim that for every $\sigma \in \Omega^{(x,a)}$,

$$W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot), Q(\sigma,\cdot)) \le \frac{1}{n}. \tag{3.22}$$

To see this, let σ_1 and σ_2 be the configurations after one step of P and Q respectively when starting from σ. We can couple σ_1 and σ_2 by picking the same vertex to update in the Glauber dynamics. If the picked vertex is not x, then we can make $\sigma_1 = \sigma_2$; meanwhile, if x is picked, which happens with probability $1/n$, then $d_{\mathrm{H}}(\sigma_1,\sigma_2) \le 1$ where the discrepancy is caused by the pinning $\sigma_x = a$. Therefore, the 1-Wasserstein distance between σ_1 and σ_2 is upper bounded by $1/n$; this justifies our claim. Combining $L_{d_{\mathrm{H}}}(f) \le 2$ and (3.22), we obtain from (3.21) that $S(x,a) \le \frac{2}{(1-\kappa)n}$ for each (x,a); consequently, $\lambda_1(J) \le \frac{2}{(1-\kappa)n}$. The same argument holds for μ^τ under any pinning τ as well, and spectral independence then follows. □

3.2 Contraction for Glauber dynamics and general metrics In this section, we generalize the Hamming metric assumption in Theorem 3.1 to any weighted Hamming metric or any metric equivalent to Hamming, which establishes Theorem 1.6. We restate it here for convenience.

THEOREM 3.2.

(1) If μ is κ-contractive w.r.t. the Glauber dynamics and an arbitrary w-weighted Hamming metric, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \le 1 - \frac{\epsilon}{n}$, then $\eta \le \frac{2}{\epsilon}$.

(2) If the metric in (1) is not a weighted Hamming metric but instead an arbitrary γ-equivalent metric, then $\eta = \frac{2\gamma^2}{(1-\kappa)n}$. In particular, if $\kappa \le 1 - \frac{\epsilon}{n}$, then $\eta \le \frac{2\gamma^2}{\epsilon}$.

We prove the two cases of Theorem 1.6 separately. We first consider the weighted Hamming metric. Recall that for a positive weight function $w : V \to \mathbb{R}_+$, the w-weighted Hamming metric $d = d_w$ is given by

$$d_w(\sigma,\tau) = \sum_{x\in V} w(x)\mathbf{1}\{\sigma_x \ne \tau_x\} \text{ for } \sigma,\tau \in \Omega.$$

In particular, if $w(x) = 1$ for all x then d is the usual Hamming metric.

Unfortunately, the proof of Theorem 3.1 does not work directly in this scenario. The reason is that the right-hand side of (3.21), with d_{H} replaced by $d = d_w$ now, can be as large as $O(w_{\max}/w_{\min})$ (more specifically, $L_d(f) = O(1/w_{\min})$ and $W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) = O(w_{\max})$), which can be unbounded since we are not making any assumption on w. To deal with this, we need to take the vertex weights into account when defining the function f and, more importantly, when defining the absolute sum of influences $S(x,a)$.

Proof of Theorem 1.6(1). For ease of notation we may assume that there is no pinning; the proof remains the same with an arbitrary pinning τ. For fixed $(x,a) \in \mathcal{X}$, we define the w-weighted sum of absolute influences given by

$$S_w(x,a) = \sum_{(y,a')\in\mathcal{X}} w(y)\,|J(x,a;y,a')|. \tag{3.23}$$

Such weighted sums were considered in [16, Lemma 22] to deduce spectral independence. We claim that if $S_w(x,a) \le \eta\, w(x)$ for all $(x,a) \in \mathcal{X}$ for some $\eta > 0$, then $\lambda_1(J) \le \eta$. To see this, let $\tilde{w} \in \mathbb{R}_+^{|\mathcal{X}|}$ with $\tilde{w}(x,a) = w(x)$ and let $W = \mathrm{diag}(\tilde{w})$; the assumption of the claim then implies that $\|W^{-1}JW\|_\infty \le \eta$ and thus $\lambda_1(J) = \lambda_1(W^{-1}JW) \le \eta$. Therefore, it suffices to upper bound the ratio $S_w(x,a)/w(x)$.

Let $\nu = \mu^{(x,a)} = \mu(\cdot \mid \sigma_x = a)$ be the conditional distribution with pinning $\sigma_x = a$, and define

$$f_w(\sigma) = \sum_{(y,a') \in \mathcal{X}} w(y)\, t(x,a;y,a')\, \mathbf{1}_{\{\sigma_y = a'\}} \tag{3.24}$$

where $t(x,a;y,a') = \mathrm{sgn}(J(x,a;y,a'))$. Observe that $L_d(f_w) \le 2$ and

$$S_w(x,a) = \mathbb{E}_\nu f_w - \mathbb{E}_\mu f_w.$$

It then follows from Lemma 3.1 that

$$S_w(x,a) \le \frac{2}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \right]$$

where P, Q are the Glauber dynamics for μ, ν respectively. For every $\sigma \in \Omega^{(x,a)}$ we have

$$W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \le \frac{w(x)}{n},$$

since if we couple the configurations σ_1, σ_2 after one step of P, Q respectively by picking the same vertex to update, then $d(\sigma_1,\sigma_2) = w(x)$ only when the site x is picked, and $\sigma_1 = \sigma_2$ otherwise. Therefore, we get $S_w(x,a) \le \frac{2w(x)}{(1-\kappa)n}$ for every $(x,a) \in \mathcal{X}$, implying that $\lambda_1(J) \le \frac{2}{(1-\kappa)n}$. The same argument works for μ^τ under any pinning τ as well, which establishes spectral independence. □

Next we consider the second part of Theorem 1.6. Recall that a metric d on Ω is said to be γ-equivalent (to the Hamming metric) for some $\gamma > 1$ if for all $\sigma, \tau \in \Omega$

$$\frac{1}{\gamma} d_{\mathrm{H}}(\sigma,\tau) \le d(\sigma,\tau) \le \gamma d_{\mathrm{H}}(\sigma,\tau).$$

To prove the second part, we follow the proof approach for Theorem 3.1, and in particular the right-hand side of (3.25) below (analogous to (3.21)) can be upper bounded using the γ-equivalence.

Proof of Theorem 1.6(2). For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 3.1 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le \frac{L_d(f)}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \right] \tag{3.25}$$

where $S(x,a)$ and f are defined by (3.17), (3.18) respectively, and P, Q are the Glauber dynamics for μ and $\nu = \mu^{(x,a)} = \mu(\cdot \mid \sigma_x = a)$ respectively. Since d is γ-equivalent, for all $\sigma, \tau \in \Omega$ we have

$$|f(\sigma) - f(\tau)| \le 2 d_{\mathrm{H}}(\sigma,\tau) \le 2\gamma d(\sigma,\tau);$$

this shows $L_d(f) \le 2\gamma$. Meanwhile, by the definition of 1-Wasserstein distance for every $\sigma \in \Omega^{(x,a)}$ we have

$$\begin{aligned} W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) &= \inf \{ \mathbb{E}_\pi[d(\sigma,\tau)] \mid \pi \in \mathcal{C}(P(\sigma,\cdot), Q(\sigma,\cdot)) \} \\ &\le \gamma \inf \{ \mathbb{E}_\pi[d_{\mathrm{H}}(\sigma,\tau)] \mid \pi \in \mathcal{C}(P(\sigma,\cdot), Q(\sigma,\cdot)) \} = \gamma W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot), Q(\sigma,\cdot)) \le \frac{\gamma}{n} \end{aligned}$$

where the last inequality is (3.22). Thus, we obtain from (3.25) that $S(x,a) \le \frac{2\gamma^2}{(1-\kappa)n}$. The rest of the proof is the same as Theorem 3.1. □

3.2.1 Application: Dobrushin uniqueness condition As an application of Theorem 1.6, we show that the Dobrushin uniqueness condition, as well as its generalizations [27, 20], implies spectral independence. Recall that the Dobrushin dependency matrix R is a $|V| \times |V|$ matrix defined as $R(x,x) = 0$ and

$$R(x,y) = \max\left\{d_{\mathrm{TV}}\left(\mu_y(\cdot \mid \sigma), \mu_y(\cdot \mid \tau)\right) : (\sigma,\tau) \in \mathcal{S}_{x,y}\right\} \text{ for } x \neq y$$

where $\mathcal{S}_{x,y}$ is the set of pairs of configurations on $V \setminus \{y\}$ that differ at most at x. Denote the spectral radius of a square matrix M by $\varrho(M)$. If M is nonnegative, then $\varrho(M)$ is an eigenvalue of M by the Perron-Frobenius theorem. We prove Theorem 1.8 from the introduction.

THEOREM 3.3. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \leq 1-\epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

REMARK 3.2. *If $\|R\|_\infty < 1$, then the Glauber dynamics mixes rapidly by a simple application of the path coupling method of Bubley and Dyer [10]. The same is true under the Dobrushin uniqueness condition, i.e., when $\|R\|_1 < 1$. Hayes [27] generalized the condition to the spectral norm $\|R\|_2 < 1$. Dyer, Goldberg, and Jerrum [20] further improved it to $\|R\| < 1$ for any matrix norm (where the mixing time depends logarithmly on the norm of the all-one matrix). Our condition $\varrho(R) < 1$ in Theorem 1.8 is technically better than previous works since for a nonnegative matrix R one has $\varrho(R) \leq \|R\|$ for any matrix norm, and the inequality can be strict for all norms when R is not irreducible; see [20] for related discussions. Finally, we point out that if R is symmetric then $\varrho(R) = \|R\|_2$.*

It is known that the Glauber dynamics is contractive for some weighted Hamming metric if the weight vector satisfies a spectral condition related to R.

LEMMA 3.2. ([20, LEMMA 20]) *If $w \in \mathbb{R}_+^V$ is a positive vector such that $Rw \leq (1-\epsilon)w$ entrywisely, then μ is $(1-\epsilon/n)$-contractive w.r.t. the Glauber dynamics and the w-weighted Hamming metric $d = d_w$.*

The following fact about nonnegative matrices is helpful.

LEMMA 3.3. ([36, EXAMPLE 7.10.2]) *If $M, N \in \mathbb{R}_+^{n\times n}$ are two nonnegative square matrices such that $M \leq N$ entrywisely, then $\varrho(M) \leq \varrho(N)$.*

We give below the proof of Theorem 1.8.

Proof of Theorem 1.8. Consider first the case that there is no pinning. If the Dobrushin dependency matrix R is irreducible, then the right principal eigenvector w associated with the eigenvalue $\varrho(R)$ satisfies $Rw = \varrho(R)w \leq (1-\epsilon)w$ and $w > 0$ by the Perron-Frobenius theorem. Hence, Lemma 3.2 and (the proof of) Theorem 1.6(1) immediately yield $\lambda_1(J) \leq 2/\epsilon$. However, if R is reducible, we cannot use the principal eigenvector directly since it may have zero entries. We instead consider the matrix $R_\delta = R + \delta O$ where O is the all-one matrix and $\delta > 0$ is a tiny constant. Let w_δ be the right principal eigenvector of R_δ associated with the eigenvalue $\varrho(R_\delta)$. Since R_δ is irreducible, $w_\delta > 0$ by the Perron-Frobenius theorem. Moreover, $Rw_\delta \leq R_\delta w_\delta = \varrho(R_\delta)w_\delta$. Since $\lim_{\delta\to 0} R_\delta = R$, we have $\lim_{\delta\to 0}\varrho(R_\delta) = \varrho(R)$; see, e.g., Remark 3.4 in [1]. Thus, $\varrho(R_\delta) < 1$ for sufficiently small δ. Then by Lemma 3.2 and Theorem 1.6(1), for δ small enough, we have $\lambda_1(J) \leq 2/(1-\varrho(R_\delta))$. Taking $\delta \to 0$ and using the assumption that $\varrho(R) \leq 1-\epsilon$, we obtain $\lambda_1(J) \leq 2/\epsilon$.

Next, consider the conditional measure μ^τ with a pinning τ on a subset $U \subset V$. Let R^τ be the Dobrushin dependency matrix for μ^τ; note that by definition $R^\tau(x,y) = 0$ if $x \in U$ or $y \in U$, and $R^\tau(x,y) \leq R(x,y)$ for all $x, y \in V$. We deduce from Lemma 3.3 that $\varrho(R^\tau) \leq \varrho(R) \leq 1-\epsilon$ and thus this is reduced to the no-pinning case. Therefore, we get $\lambda_1(J^\tau) \leq 2/\epsilon$ for all τ and spectral independence then follows. □

3.3 Contraction for general Markov chains and general metrics In this section, we generalize Theorem 3.1 to arbitrary "local" Markov chains and arbitrary metrics close to the Hamming metric. In particular, we prove Theorem 1.7.

Consider a collection of Markov chains $\mathcal{P} = \{P^\tau : \tau \in \mathcal{T}\}$ associated with μ, where each P^τ is a Markov chain on Ω^τ with stationary distribution μ^τ. Intuitively, one can think of $\mathcal{P}$ as the same dynamics applied to

all conditional distributions μ^τ; for example, $\mathcal{P}$ can be the collection of Glauber dynamics for all μ^τ's. We are particularly interested in local dynamics; these are Markov chains that make local updates on the configuration in each step, e.g., Glauber dynamics for spin systems or flip dynamics for colorings. Alternatively, we can describe local dynamics as those insensitive to pinnings; that is, if the dynamics is applied to both μ and $\mu^{(x,a)}$ with a pinning $\sigma_x = a$, then with high probability there is no difference in the two chains or the discrepancy caused by the pinning will not propagate. This motivates the following definition.

DEFINITION 3.1. *We say a collection $\mathcal{P}$ of Markov chains associated with μ is* Φ-local *if for any two adjacent pinnings $\tau \in \mathcal{T}$ and $\tau' = \tau \cup (x,a)$ where $(x,a) \in \mathcal{X}^\tau$ (i.e., τ' combines τ and the pinning $\sigma_x = a$), and for all $\sigma \in \Omega^{\tau'}$, we have*

$$W_{1,d_{\mathrm{H}}}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \le \Phi.$$

We show that for such local dynamics contraction implies spectral independence.

THEOREM 3.4. *If μ is κ-contractive w.r.t. a Φ-local collection $\mathcal{P}$ of Markov chains and a γ-equivalent metric d for some $\kappa \in (0,1)$, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2\Phi}{1-\kappa}$.*

Proof. The proof is similar to that of Theorems 3.1 and 1.6(2). For an arbitrary pinning τ and $(x,a) \in \mathcal{X}^\tau$, we define

$$S^\tau(x,a) = \sum_{(y,a') \in \mathcal{X}^\tau} |J^\tau(x,a;y,a')| \tag{3.26}$$

and

$$f^\tau(\sigma) = \sum_{(y,a') \in \mathcal{X}^\tau} t^\tau(x,a;y,a')\, \mathbf{1}_{\{\sigma_y = a'\}} \tag{3.27}$$

where $t^\tau(x,a;y,a') = \mathrm{sgn}(J^\tau(x,a;y,a'))$; these definitions are analogous to (3.17) and (3.18) with pinning τ. Let $\tau' = \tau \cup (x,a)$. Then we deduce from Lemma 3.1 that

$$S^\tau(x,a) = \mathbb{E}_{\mu^{\tau'}} f^\tau - \mathbb{E}_{\mu^\tau} f^\tau \le \frac{L_d(f^\tau)}{1-\kappa}\, \mathbb{E}_{\mu^{\tau'}} \left[W_{1,d}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \right].$$

As shown in the proof of Theorem 1.6(2), since d is γ-equivalent to the Hamming metric we have $L_d(f^\tau) \le \gamma L_{d_{\mathrm{H}}}(f^\tau) \le 2\gamma$ and for all $\sigma \in \Omega^{\tau'}$ we have

$$W_{1,d}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \le \gamma W_{1,d_{\mathrm{H}}}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \le \gamma\Phi$$

using the Φ-locality of $\mathcal{P}$. Therefore, we obtain that $S^\tau(x,a) \le \frac{2\gamma^2\Phi}{1-\kappa}$ for all $(x,a) \in \mathcal{X}^\tau$. This yields $\lambda_1(J^\tau) \le \frac{2\gamma^2\Phi}{1-\kappa}$ and spectral independence follows. ∎

To better understand local dynamics, we consider a very general type of Markov chains which we call *select-update dynamics*; examples include the Glauber dynamics, heat-bath block dynamics, and flip dynamics. Let $\mathcal{B}$ be a collection of blocks associated with the select-update dynamics and fix some pinning τ. Given the current configuration $\sigma^t \in \Omega^\tau$, the next configuration σ^{t+1} is generated as follows:

1. SELECT: Select a block $B \in \mathcal{B}$ from some distribution p_t over $\mathcal{B}$;
2. UPDATE: Resample the configuration on B from some distribution ν_B^t.

We try to make weakest assumptions on the selection rule p_t and the update rule ν_B^t: the selection distribution p_t is allowed to depend on the current configuration σ^t but is independent of the pinning τ, and the update distribution ν_B^t is allowed to depend on the whole current configuration σ^t and the part of the pinning τ contained in B. In particular, the heat-bath block dynamics is a special case of the select-update dynamics: the selection rule $p_t = \alpha$ is a fixed distribution over $\mathcal{B}$ and the update rule ν_B^t is the marginal distribution on B conditioned on σ^t outside B and the pinning τ in B.

REMARK 3.3. *The assumption that the selection rule p_t is independent of the pinning τ is not necessary, but it is helpful for stating and proving our theorems and does not weaken our results. Roughly speaking, we only require that the collection of the select-update dynamics is the same dynamics applied to all μ^τ's, and the selection rule p_t can be conditioned on containing at least one unpinned vertex. See the discussions in Remark 3.1 for the Glauber dynamics.*

We write $\mathcal{P}_\mathcal{B}$ for a collection of select-update dynamics associated with μ. Denote the maximum block size of $\mathcal{B}$ by

$$M = \max_{B\in\mathcal{B}} |B|,$$

and the maximum probability of a vertex being selected in Step 1 by

$$D = \max_{p_t} \max_{x} \sum_{B\in\mathcal{B}: x\in B} p_t(B), \tag{3.28}$$

where we maximize over all selection rules p_t that can occur. We can show that the select-update dynamics $\mathcal{P}_\mathcal{B}$ is Φ-local with $\Phi = DM$; using this and Theorem 3.4 we establish Theorem 1.7, which we restate here for convenience.

THEOREM 3.5. *If μ is κ-contractive w.r.t. arbitrary select-update dynamics and an arbitrary γ-equivalent metric, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2 DM}{1-\kappa}$.*

Proof. It suffices to show that the select-update dynamics $\mathcal{P}_\mathcal{B}$ is Φ-local with $\Phi = DM$; the theorem would then follows immediately from Theorem 3.4. Consider two adjacent pinnings τ and $\tau' = \tau \cup (x, a)$ where $(x, a) \in \mathcal{X}^\tau$. For $\sigma \in \Omega^{\tau'}$, let σ_1 and σ_2 be the two configurations obtained from σ after one step of P^τ and $P^{\tau'}$ respectively. We couple σ_1 and σ_2 by picking the same block $B \in \mathcal{B}$ in Step 1 of the select-update dynamics. If $x \notin B$, then we have $\sigma_1 = \sigma_2$. Meanwhile, if $x \in B$, which happens with probability at most D, we have $d_{\mathrm{H}}(\sigma_1, \sigma_2) \le |B| \le M$. Therefore,

$$W_{1,d_{\mathrm{H}}}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \le DM.$$

This establishes the (DM)-locality for $\mathcal{P}_\mathcal{B}$. ∎

REMARK 3.4. *If we further assume that in Step 2 the select-update dynamics resamples a block independently for each of its components (i.e., the update rule ν_B^t is a product distribution over all components of the induced subgraph $G[B]$), then in Theorem 1.7 the maximum block size M can be replaced by the maximum component size of all blocks.*

3.3.1 Application: flip dynamics for colorings In this section we establish spectral independence for colorings utilizing Theorem 1.7.

THEOREM 3.6. *Let $\epsilon_0 \approx 10^{-5} > 0$ be a fixed constant. Let $\Delta, q \ge 3$ be integers and $q > (\frac{11}{6} - \epsilon_0)\Delta$. Then there exists $\eta = \eta(\Delta, q) > 0$ such that the following holds.*

Let μ be the uniform distribution over all proper q-colorings of a graph $G = (V, E)$ of maximum degree at most Δ. Then μ is spectrally independent with constant η.

To apply Theorem 1.7, we need a contractive Markov chain for sampling colorings of a graph. Vigoda considered the *flip dynamics* [44] and showed that it is contractive for the Hamming metric when the number of colors $q > \frac{11}{6}\Delta$. Recently, [14] improved the bound to $q > (\frac{11}{6} - \epsilon_0)\Delta$ for a fixed tiny constant $\epsilon_0 \approx 10^{-5}$, using variable-length coupling or an alternative metric. Our result on spectral independence builds upon contraction results for the flip dynamics.

We first describe the flip dynamics. Let Ω be the set of all proper q-colorings of G. Fix a pinning τ on $U \subset V$. For a coloring $\sigma \in \Omega$, a vertex $x \in V$, and a color $a \in [q]$, denote by $L_\sigma(x, a)$ the bicolored component containing x with colors a and σ_x; that is, the set of all vertices which can be reached from x through an alternating (σ_x, a)-colored path. Given the coloring σ^t at time t, the flip dynamics with pinning τ generates the next coloring σ^{t+1} as follows:

1. Pick a vertex $x \in V$ u.a.r. and a color $a \in [q]$ u.a.r.;
2. If $L_{\sigma^t}(x,a)$ contains a pinned vertex (i.e., $L_{\sigma^t}(x,a) \cap U \neq \emptyset$), then $\sigma^{t+1} = \sigma^t$;
3. If all vertices in $L_{\sigma^t}(x,a)$ are free (i.e., $L_{\sigma^t}(x,a) \cap U = \emptyset$), then flip the two colors of $L_{\sigma^t}(x,a)$ with probability p_s/s where $s = |L_{\sigma^t}(x,a)|$.

The flip dynamics is specified by the flip parameters $\{p_s\}_{s=1}^{\infty}$. In [44] and the recent improvement [14], the flip parameters are chosen in such a way that $p_s = 0$ for all $s \geq 7$; i.e., in each step at most six vertices change their colors. We set the flip parameters as in Observation 5.1 from [14], where the authors established contraction of the flip dynamics using the path coupling method.

LEMMA 3.4. ([14]) *Under the assumptions of Theorem 3.6, there exists a constant $\epsilon = \epsilon(\Delta, q) > 0$ and a 2-equivalent metric d such that μ is $(1-\epsilon/n)$-contractive w.r.t. the flip dynamics and the metric d.*

We remark that the pinning τ induces a list coloring instance where each unpinned vertex has a color list to choose its color from, and the results of [14] generalize naturally to list colorings. Also, in this paper we assume that the flip dynamics may pick a pinned vertex and stay at the current coloring. This does not weaken our results since we only consider the flip dynamics for analysis rather than actually running it; see Remark 3.1 addressing the same issue for the Glauber dynamics and also Remark 3.3 for general select-update dynamics.

We give below the proof of Theorem 3.6.

Proof of Theorem 3.6. Observe that the flip dynamics belongs to the class of select-update dynamics, where the associated $\mathcal{B}$ is the collection of connected subsets of vertices. Since the flip parameters satisfy $p_s > 0$ only for $s \leq 6$, we have $M \leq 6$. Moreover, we have $D \leq \Delta^6/n$ since a vertex x is in the selected bicolored component $L_{\sigma^t}(y,a)$ only if $\mathrm{dist}(x,y) \leq 5$, which happens with probability at most Δ^6/n. The theorem then follows from Lemma 3.4 and Theorem 1.7. □

We conclude here with the proof of Theorem 1.1.

Proof of Theorem 1.1. By Theorem 3.6 the uniform distribution μ of proper colorings is spectrally independent. Then the results follows immediately from Theorem 1.4. □

3.3.2 Application: block dynamics for Potts model Here we apply Theorems 3.1 and 1.7 to the ferromagnetic Potts model to establish spectral independence.

THEOREM 3.7. *Let $\Delta \geq 3$ and $q \geq 2$ be integers. Let μ be the Gibbs distribution of the q-state ferromagnetic Potts model with inverse temperature parameter β on a graph $G = (V,E)$ of maximum degree at most Δ. Then, the following holds:*

1. *If $\beta < \max\left\{\frac{2}{\Delta}, \frac{1}{\Delta}\ln(\frac{q-1}{\Delta})\right\}$, then μ is spectrally independent with constant $\eta = \eta(\beta,\Delta)$.*
2. *For any $\delta > 0$ there exists $c = c(\delta,\Delta) > 0$ such that, if $\beta \leq \frac{\ln q - c}{\Delta - 1 + \delta}$ then μ is spectrally independent with constant $\eta = \eta(\delta,\beta,\Delta)$.*

To prove this theorem, we need the following results from [43] and [8] regarding the contraction of the Glauber dynamics and of the heat-bath block dynamics with a specific choice of blocks.

LEMMA 3.5. ([43, COROLLARY 2.14] & [8, PROPOSITION 2.2]) *Under the assumptions in Part 1 of Theorem 3.7, there exists a constant $\epsilon = \epsilon(\beta,\Delta)$ such that μ is $(1-\frac{\epsilon}{n})$-contractive w.r.t. the Glauber dynamics and the Hamming metric.*

LEMMA 3.6. ([8, THEOREM 2.7]) *Under the assumptions in Part 2 of Theorem 3.7, there exists a collection of blocks $\mathcal{B} = \{B_x\}_{x\in V}$ satisfying $x \in B_x$, $|B_x| = O(1/\delta)$ and $G[B_x]$ connected for all x, such that μ is $(1-\frac{1}{2n})$-contractive w.r.t. the α-weighted heat-bath block dynamics for $\mathcal{B}$ and the Hamming metric, where α is the uniform distribution over $\mathcal{B}$.*

REMARK 3.5. *To be more precise, [8] shows that the conclusion of Lemma 3.6 is true when β, q, and the maximum block size $M = \max_{x\in V} |B_x|$ satisfy*

$$\beta\left(\Delta - 1 + \frac{1}{M}\right) + 3M(\ln\Delta + \ln M) \le \ln q. \tag{3.29}$$

Thus, for any $\delta > 0$, by taking $M = \lceil \delta^{-1} \rceil$ and $c = 3M(\ln\Delta + \ln M)$, our assumption $\beta \le \frac{\ln q - c}{\Delta - 1 + \delta}$ in Part 2 of Theorem 3.7 implies (3.29). Moreover, if we take, say, $M \approx \sqrt{\ln q}$ (namely, $\delta \approx 1/\sqrt{\ln q}$), then $c = o(\ln q)$ and our assumption becomes $\beta \le (1 - o(1))\frac{\ln q}{\Delta - 1}$ where $o(1)$ tends to 0 as $q \to \infty$; this gives the bound β_1 in Theorem 1.2 from the introduction.

Theorem 3.7 is an immediate consequence of Lemmas 3.5, 3.6 and the results proved in this section.

Proof of Theorem 3.7. Part 1 follows directly from Lemma 3.5 and Theorem 3.1. For Part 2, we note that the block dynamics from Lemma 3.6 corresponds to a select-update dynamics with $M = O(1/\delta)$ and $D = \Delta^{O(1/\delta)}/n$; the reason of the latter is that each x is in at most $\Delta^{O(M)}$ blocks. The theorem then follows from Lemma 3.6 and Theorem 1.7. □

We end this section with the proof of Theorem 1.2.

Proof of Theorem 1.2. For Ising model, spectral independence is known in the whole uniqueness region [16]. For Potts model, Theorem 3.7 establishes spectral independence in the corresponding parameter regimes. The theorem then follows from Theorems 1.4 and 1.5. □

References

[1] A. Alexanderian. On continuous dependence of roots of polynomials on coefficients. Technical notes, 2013.

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

[3] R. Bauerschmidt and T. Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019.

[4] A. Blanca, P. Caputo, Z. Chen, D. Parisi, D. Štefankovič, and E. Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. *Preprint, arXiv:2103.07459*, 2021.

[5] A. Blanca, P. Caputo, D. Parisi, A. Sinclair, and E. Vigoda. Entropy decay in the Swendsen-Wang dynamics. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021.

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

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

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

[9] G. Bresler and D. Nagaraj. Stein's method for stationary distributions of Markov chains and application to Ising models. *The Annals of Applied Probability*, 29(5):3230–3265, 2019.

[10] R. Bubley and M. E. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 223–231, 1997.

[11] P. Caputo, G. Menz, and P. Tetali. Approximate tensorization of entropy at high temperature. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, 24(4):691–716, 2015.

[12] P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, pages 1–26, 2021.

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

[14] S. Chen, M. Delcourt, A. Moitra, G. Perarnau, and L. Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2216–2234, 2019.

[15] Z. Chen, A. Galanis, D. Štefankovič, and E. Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557, 2021.
[16] Z. Chen, K. Liu, and E. Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, 2020.
[17] Z. Chen, K. Liu, and E. Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021.
[18] M. Cryan, H. Guo, and G. Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *The Annals of Probability*, 49(1):506–525, 2021.
[19] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 9(3):695–750, 1996.
[20] M. Dyer, L. A. Goldberg, and M. Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71–107, 2009.
[21] R. G. Edwards and A. D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical review D*, 38(6):2009, 1988.
[22] R. Eldan, J. R. Lee, and J. Lehec. Transport-entropy inequalities and curvature in discrete-space Markov chains. In *A Journey Through Discrete Mathematics*, pages 391–406. 2017.
[23] W. Feng, H. Guo, Y. Yin, and C. Zhang. Rapid mixing from spectral independence beyond the Boolean domain. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1558–1577, 2021.
[24] F. Götze, H. Sambale, and A. Sinulis. Higher order concentration for functions of weakly dependent random variables. *Electronic Journal of Probability*, 24, 2019.
[25] G. R. Grimmett. *The Random-Cluster Model*, volume 333. Springer-Verlag, Berlin, 2006.
[26] O. Häggström. The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104(2):231–253, 1996.
[27] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 39–46, 2006.
[28] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Annals of Applied Probability*, 17(3):931–952, 2007.
[29] J. Hermon and J. Salez. Modified log-Sobolev inequalities for strong-Rayleigh measures. *Preprint, arXiv:1902.02775*, 2019.
[30] M. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.
[31] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.
[32] K. Liu. From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk. In *APPROX-RANDOM*, 2021.
[33] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics*, pages 93–191. Springer, 1999.
[34] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. *Communications in Mathematical Physics*, 161(3):447–486, 1994.
[35] K. Marton. Logarithmic Sobolev inequalities in discrete product spaces. *Combinatorics, Probability & Computing*, 28(6):919–935, 2019.
[36] C. D. Meyer. *Matrix analysis and applied linear algebra*, volume 71. SIAM, 2000.
[37] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.
[38] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.
[39] G. Reinert and N. Ross. Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. *The Annals of Applied Probability*, 29(5):3201–3229, 2019.
[40] H. Sambale and A. Sinulis. Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli*, 26(3):1863–1890, 2020.
[41] D. W. Stroock and B. Zegarlinski. The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. *Communications in Mathematical Physics*, 144(2):303–323, 1992.
[42] R. H. Swendsen and J. S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.
[43] M. Ullrich. Rapid mixing of Swendsen-Wang dynamics in two dimensions. *Dissertationes Mathematicae*, 502:1–65, 2014.
[44] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.