Electron. J. Probab. **27** (2022), article no. 142, 1–42.
ISSN: 1083-6489 https://doi.org/10.1214/22-EJP867

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.

*Supported in part by NSF grants CCF-1850443, CCF-2007287, and CCF-2007022.
†Department of Computer Science and Engineering, Penn State, University Park, PA 16801, USA.
E-mail: ablanca@cse.psu.edu
‡Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy.
E-mail: caputo@mat.uniroma3.it
§Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: zongchen@mit.edu
¶Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy.
E-mail: daniel.parisi@uniroma3.it
‖Department of Computer Science, University of Rochester, Rochester, NY 14627, USA.
E-mail: stefanko@cs.rochester.edu
**Department of Computer Science, University of California, Santa Barbara, CA 93106, USA.
E-mail: vigoda@ucsb.edu

Keywords: MCMC; mixing time; spectral independence; Swendsen-Wang; log-Sobolev.
MSC2020 subject classifications: 60J10; 82B20; 68Q87.
Submitted to EJP on December 20, 2021, final version accepted on October 17, 2022.

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 [ALO20] 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 [CP21, BCP$^+$21].

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.

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 [BD97]. 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., [DS96, Mar99, BT06].

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. Here 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 context of spin systems on bounded-degree graphs, this confirms a remarkable (and more general) conjecture of Peres and Tetali (see Conjecture 3.1 in [ELL17] and Remark 1.14). We refer to [CDPP09, EHMT17, Con22, Sal21] for further relevant works on the relations between curvature and entropy in Markov chains.

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 [HS07] 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. [BT06] 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 [Jer95] 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 [Vig00] and further improved to $q > (\frac{11}{6} - \epsilon_0)\Delta$ for some small $\epsilon_0 \approx 10^{-5} > 0$ by Chen et al. [CDM$^+$19] 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 [HS07], 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 [MS13] 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 [SW87]. 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 [BCV20], 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 [BCP$^+$21] 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 [Hay06, Ull14, BGP16] 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, [BGP16] showed $\mathrm{poly}(n)$ 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 4.16 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 [Häg96] and it behaves as $\beta_u = \frac{\ln q}{\Delta-1} + O(1)$; see [BGP16]. 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 [ALO20] 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 4.2 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 1.3 (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 [ALO20] 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 [CGvV21, FGYZ21] in different ways. In this paper we use the definition from [CGvV21] which is more suitable for our applications in Section 4 for establishing spectral independence, but we could also work with the definition from [FGYZ21] 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 1.4 (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 1.5 (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 [ALO20], for any 2-spin antiferromagnetic spin system in the tree uniqueness region [CLV20], and for colorings [CGvV21, FGYZ21] it matched the best known parameter bounds using other algorithmic approaches. Moreover, recent work of Chen et al. [CLV21] shows that spectral independence implies optimal mixing of the Glauber dynamics in all of these cases as stated in the following theorem.

Theorem 1.6 ([CLV21])**.** *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 [CLV21]) to $O(1+\eta/b)$.

The key step in the proof of Theorem 1.6 is the implication

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

Approximate tensorization of entropy for a Gibbs distribution μ, as defined in [CMT15], says that there exists a constant $C \geq 1$, such that for any function $f : \Omega \to \mathbb{R}_+$,

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

Here $\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$ 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 for the detailed definitions. The inequality (1.3) generalizes the well known tensorization property of product measures: if μ is product, then (1.3) holds with $C = 1$. 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. [CMT15]. 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 [SZ92, MO94, Ces01, DPPP02].

We provide an alternative proof of some of the key steps for the implication (1.2). The analogous result in [CLV21] is proved in the framework of simplicial complexes and generalizes the result of [CGM21] for homogeneous strongly log-concave distributions; see also [HS19] for related results. Our proof, which is provided in Section 5, 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 1.6.

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

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

The notion of *general block factorization* of entropy, recently introduced in [CP21], is a generalization of the 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 [CMT15]. 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 (see Lemma 2.8 below) 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 1.7. *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, using the simple facts recalled in Lemma 2.8 below, one recovers Theorem 1.6 as a special case of the above result. As another example, for a bipartite graph, Theorem 1.7 implies $O(\log n)$ mixing time of the even-odd dynamics where at each step, according to a coin flip, all the odd sites or all the even sites are updated at once.

When α is the uniform distribution over all subsets of a given size ℓ, we refer to (1.5) as the *ℓ-uniform block factorization of entropy* or ℓ-UBF for short. In [CLV21], an important step in the proof of Theorem 1.6 is establishing ℓ-UBF with $\ell \sim \theta n$ for some $\theta \in (0, 1)$. To prove Theorem 1.7 for arbitrary blocks we establish that ℓ-UBF implies general block factorization of entropy; see Theorem 1.16 for a detailed statement and Figure 1 for a high-level overview.

Recent work of Blanca et al. [BCP$^+$21] 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 [BCP$^+$21] and using our general result in Theorem 1.7, 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 1.8. *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 1.9 (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 1.10. *(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 that a κ-contractive coupling for the Hamming distance immediately implies $O(n \log n)$ mixing time of the Glauber dynamics (see, e.g., [BD97, LP17]). 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 [BD97], 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 1.10 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 4.3.1.

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 (4.12) for the precise definition of D.

Theorem 1.11. *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 1.11 generalizes Theorem 1.10(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 4.10. See also Theorem 4.8 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 1.11 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 1.10 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 1.12 (Dobrushin uniqueness condition)**.** *The* Dobrushin dependency (or influence) matrix $R \in \mathbb{R}^{V\times V}$ *is defined by* $R(x,x) = 0$ *and*

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

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 1.10 we prove that the Dobrushin uniqueness condition implies spectral independence. Moreover, our result holds under generalizations of the Dobrushin uniqueness condition. Hayes [Hay06] generalized it to the following spectral condition: if $\|R\|_2 \le 1-\epsilon$ 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. [DGJ09] 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 4.4 for a more detailed discussion.

Theorem 1.13. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \le 1-\epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

Previously, Marton [Mar19] (see also [GSS19, SS20]) showed that the spectral condition in Theorem 1.13 implies approximate tensorization of entropy and thus optimal

bounds on the modified log-Sobolev constant for the Glauber dynamics. However, the approach in these works does not imply block factorization of entropy as in our case.

Remark 1.14. 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 [Oll09]. Combining Theorem 1.10 with Theorem 1.7 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 1.10 with its generalization Theorem 1.11 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. [ELL17].

1.5 Proof overviews for main results

In this section, we sketch the proof of our key technical results. We begin in Section 1.5.1 with an overview of our proof that spectral independence implies optimal bounds for arbitrary block dynamics and the Swendsen-Wang dynamics (namely, Theorems 1.7 and 1.8). In Section 1.5.2 we highlight the proofs of Theorems 1.10, and 1.11 that a contractive coupling for an arbitrary local dynamics implies spectral independence.

1.5.1 Optimal mixing under spectral independence: Theorems 1.7 and 1.8

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 Lemma 2.8 for more details. Previous results show that spectral independence implies ℓ-uniform block factorization (ℓ-UBF) with $\ell = \lceil \theta n \rceil$ for any fixed $\theta \in (0,1)$; see [CLV21] and Theorem 5.1. Note, ℓ-UBF refers to the block factorization where the weights α are uniform over all subsets of size ℓ; see Definition 1.15 below. The key step in the proof of Theorem 1.7 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 1.15 (Uniform Block Factorization)**.** *We say that the spin system μ satisfies the* ℓ-uniform block factorization (ℓ-UBF) of entropy *with constant C_{ubf} if for all $f : \Omega \to \mathbb{R}_+$*

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

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

Theorem 1.16. *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 $C_{\mathrm{gbf}} = C_{\mathrm{ubf}} \times O\left((\theta\, b^2)^{-1} \log(1/b)\Delta^3\right)$.*

This and the already known $\lceil \theta n \rceil$-UBF implies Theorem 1.7 from the introduction:

Proof of Theorem 1.7. For a spin system which is η-spectrally independent and also 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). Then, taking $\theta = \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, Theorem 1.16 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 $C_{\text{gbf}} = \left(\frac{2}{b}\right)^{O\left(\Delta(1+\frac{\eta}{b})\right)}$. □

Hence, the key novelty in the proof of Theorem 1.7 is Theorem 1.16. To establish Theorem 1.16 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 1.16 follows immediately from the following factorization statements.

Lemma 1.17. *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,*

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

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 1.18. *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 are provided in Section 3. The main idea behind the proof of Lemma 1.17 can be roughly explained as follows. If the sets S in (1.6) were all independent sets, then a suitable decomposition of the entropy functional would imply the desired conclusion. Using a tensorization argument from [CP21], the same conclusion would continue to hold if S only contained connected components of bounded size. However, even if θ is small, a uniformly random set S with $|S| = \lceil \theta n \rceil$ is likely to have components of size $\Theta(\log n)$. On the other hand, locally the expected component size is bounded if θ is sufficiently small. The challenge in obtaining optimal bounds is thus to use the expected local component size instead of the maximum component size. To achieve this we combine ideas from [CP21] and [CLV21] together with a new conditioning argument. The proof of Lemma 1.18 is simpler, and 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 [CP21]).

Finally, to prove Theorem 1.8, that is the optimal mixing results for the SW dynamics, our strategy is based on establishing the *spin/edge factorization* of entropy, a notion introduced in [BCP$^+$21], for the "joint" Edwards-Sokal coupling; see [ES88, Gri06] and Section 6 below. The spin/edge factorization of entropy was shown in [BCP$^+$21, Lemma 1.8] to imply $O(\log n)$ mixing of the SW dynamics on any graph. To prove Theorem 1.8, we show that k-partite factorization of entropy for μ implies spin/edge factorization of entropy. This requires a nontrivial adaptation of the corresponding result established in [BCP$^+$21] in the special case of bipartite graphs. The proof of Theorem 1.8 is provided in Section 6.

1.5.2 Spectral independence via contractivity: Theorems 1.10 and 1.11

Here we outline our proofs of Theorems 1.10 and 1.11. We establish spectral independence by showing that the maximum absolute row sum of the ALO influence matrix is

bounded. Consider the case without pinnings for simplicity. We would like to upper bound, for each $(x,a) \in \mathcal{X}$, the quantity

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

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. The first step is to 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, it follows that $S(x,a) \leq 2W_{1,d_{\mathrm{H}}}(\nu,\mu)$ where $W_{1,d_{\mathrm{H}}}(\nu,\mu)$ represents the 1-Wasserstein distance; we refer to Section 4.1 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.

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 [BN19, RR19] to bound the Wasserstein distance of two distributions; see Lemma 4.3. 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 [BN19, RR19] were specialized for the binary product space and the Glauber dynamics. Here, we establish our Lemma 4.3 for any finite state space and any Markov chain. This result is of independent interest and may find applications in other problems.

We point out that after the first version of the present paper appeared, Kuikui Liu posted the article [Liu21] where he independently established similar results to Theorem 1.11, which also implied Theorem 1.1 for the Glauber dynamics. In [Liu21] the author concluded a version of Theorem 1.13 as well, but required the stronger assumption that the row sum of the Dobrushin dependency matrix is bounded. Using our Theorem 1.10 we only require a bound on the spectral radius which is the weakest assumption of this type; see Remark 4.4.

Paper organization The organization of the paper is demonstrated in Figure 1. After giving preliminaries in Section 2, in Section 3 we prove that uniform block factorization implies general block factorization of entropy. In Section 4, we establish spectral independence if the distribution admits a contractive Markov chain. In Section 5, we reformulate the result of [CLV21] showing that spectral independence implies uniform block factorization; our new proof avoids abstract simplicial complexes and gives a slightly better constant. We also show in Section 5 that spectral independence implies approximate subadditivity of entropy, see Theorem 5.1. Finally, in Section 6 we show optimal mixing and optimal entropy decay of the Swendsen-Wang dynamics if k-partite factorization holds, which can be in turn deduced from spectral independence.

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 $[q] = \{1, \ldots, q\}$. 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 $A_{xy} \in \mathbb{R}_+^{q \times q}$,

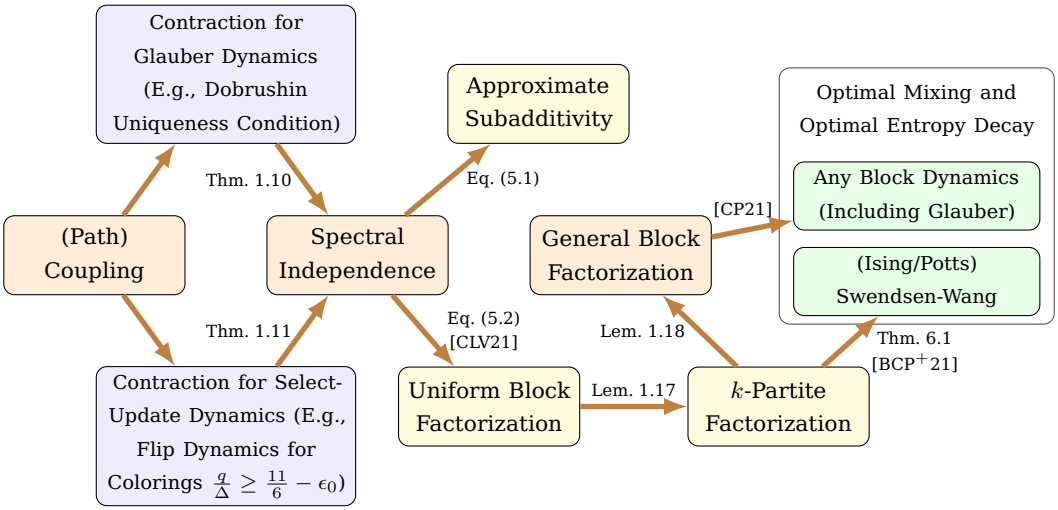

Figure 1: Organization of the paper

$\{x, y\} \in E \cup \partial E$, 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 feasible 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 $\tau \in \Omega_U$, let μ^τ denote the conditional Gibbs distribution $\mu(\cdot \,|\, \sigma_U = \tau)$. We also write $\mu_\Lambda^\tau = \mu^\tau$ if $\tau \in \Omega_{V \setminus \Lambda}$ and use the notation $\mu_\Lambda : \ \Omega_{V \setminus \Lambda} \ni \tau \mapsto \mu_\Lambda^\tau$ for the associated mapping. Following a standard convention, with slight abuse of notation we sometimes consider μ_Λ as a map on the whole set Ω_V, that is $\mu_\Lambda : \ \Omega_V \ni \sigma \mapsto \mu_\Lambda^\sigma$ in such a way that $\mu_\Lambda^\sigma = \mu_\Lambda^{\sigma'}$ for all $\sigma, \sigma' \in \Omega_V$ which coincide on $V \setminus \Lambda$.

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 $\Omega_W^\tau = \{\varphi \in [q]^W : \exists \sigma \in \Omega^\tau, \sigma_W = \varphi\}$ for $W \subset V \setminus U$ and $\mathcal{X}^\tau = \{(x, a) : x \in V \setminus U, a \in \Omega_x^\tau\}$. 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* $T_{\mathrm{mix}}(P)$ of the chain is defined as

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

where $\|\cdot\|_{\mathrm{TV}}$ 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 $\mathrm{Var}_\mu(f) = \mu[f^2] - \mu[f]^2$ 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

$$\mathrm{Ent}(f) = \mu\left[f \cdot \log\left(\frac{f}{\mu[f]} \right) \right] = \mu[f \cdot \log f] - \mu[f] \cdot \log \mu[f]. \tag{2.1}$$

When $f \geq 0$ is such that $\mu[f] = 1$, then $\mathrm{Ent}(f) = H(f\mu \,|\, \mu)$ 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

$$\mathcal{D}_P(f, g) = \langle f, (1 - P)g \rangle_\mu, \tag{2.2}$$

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

$$\mathcal{D}_P(f, g) = \frac{1}{2} \sum_{\sigma, \tau \in \Omega} \mu(\sigma) P(\sigma, \tau)(f(\sigma) - f(\tau))(g(\sigma) - g(\tau)). \tag{2.3}$$

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

$$\mathcal{D}_P(\sqrt{f}, \sqrt{f}) \geq s \, \mathrm{Ent}(f). \tag{2.4}$$

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

$$\mathcal{D}_P(f, \log f) \geq \varrho \, \mathrm{Ent}(f). \tag{2.5}$$

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

$$H(\nu P \,|\, \mu) \leq (1 - \delta) H(\nu \,|\, \mu). \tag{2.6}$$

In this paper we focus on the entropy decay inequality (2.6) 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 2.2. *If (P,μ) satisfies the standard LSI with constant $s>0$ 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*

$$T_{\mathrm{mix}}(P)\le 1+\delta^{-1}[\log(8)+\log\log(1/\mu_*)]\,, \tag{2.7}$$

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

We refer to e.g. [BCP$^+$21, Section 2] for a proof. If P is positive semi-definite, and therefore (P,μ) is reversible, then one can additionally show that the standard LSI with constant s implies the discrete time relative entropy decay with rate $\delta=s$ (see [BCP$^+$21, Lemma 2.8], whose statement erroneously omitted the requirement that P is positive semi-definite).

2.3 Some basic properties of entropy

To compute the relative entropy with respect to a pinned measure μ_Λ^τ it is convenient to use the notation

$$\mathrm{Ent}_\Lambda(f)=\mu_\Lambda\left[f\log\left(f/\mu_\Lambda[f]\right)\right], \tag{2.8}$$

with the understanding that if we evaluate the left hand side at a given pinning τ on $\Lambda^c=V\setminus\Lambda$ we then evaluate the expectations in the right hand side with respect to μ_Λ^τ. To emphasize the dependence on the pinning we sometimes write $\mathrm{Ent}_\Lambda^\tau(f)$. The expectation $\mu[\mathrm{Ent}_\Lambda f]$ is obtained by averaging with respect to μ over the pinning τ on Λ^c, and satisfies

$$\mu[\mathrm{Ent}_\Lambda(f)]=\sum_{\tau\in\Omega_{\Lambda^c}}\mu(\sigma_{\Lambda^c}=\tau)\,\mathrm{Ent}_\Lambda^\tau(f)=\mu\left[f\log\left(f/\mu_\Lambda[f]\right)\right]. \tag{2.9}$$

The following lemma summarizes a standard decomposition of the relative entropy; see e.g. [CP21, Lemma 3.1] for a proof.

Lemma 2.3. *For any $\Lambda\subset V$, for any $f:\Omega\to\mathbb{R}_+$:*

$$\mathrm{Ent}(f)=\mu\left[\mathrm{Ent}_\Lambda(f)\right]+\mathrm{Ent}\left(\mu_\Lambda[f]\right). \tag{2.10}$$

More generally, for any $\Lambda_0\subset\Lambda_1\subset\cdots\subset\Lambda_w\subset V$, for any $f:\Omega\to\mathbb{R}_+$:

$$\sum_{i=1}^{w}\mu\left[\mathrm{Ent}_{\Lambda_i}(\mu_{\Lambda_{i-1}}[f])\right]=\mu\left[\mathrm{Ent}_{\Lambda_w}(\mu_{\Lambda_0}[f])\right]. \tag{2.11}$$

The following monotonicity property of the entropy functional is an immediate consequence of the previous lemma.

Lemma 2.4. *For all $A\subset B\subset V$,*

$$\mu[\mathrm{Ent}_A(f)]\le\mu[\mathrm{Ent}_B(f)]\,. \tag{2.12}$$

Next, we recall the definition of general block factorization of entropy.

Definition 2.5. *The spin system is said to satisfy the* general block factorization of entropy *with constant C if for all $f\ge 0$, for all probability distribution α over subsets of V,*

$$\delta(\alpha)\,\mathrm{Ent}\,f\le C\sum_{B\subset V}\alpha_B\,\mu[\mathrm{Ent}_B\,f], \tag{2.13}$$

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

We will often consider independent sets Λ of V, that is sets of vertices whose induced subgraph in G has no edge; in those cases, μ_Λ is a product measure $\mu_\Lambda = \otimes_{x\in\Lambda}\mu_x$ and the following lemma will be useful.

Lemma 2.6. *Fix $\Lambda \subset V$ and suppose that μ_Λ is a product measure $\mu_\Lambda = \otimes_{x\in\Lambda}\mu_x$. Then, for any distribution α over the subsets of Λ, and any $f : \Omega \to \mathbb{R}_+$:*

$$\delta(\alpha)\operatorname{Ent}_\Lambda(f) \le \sum_{B\subset\Lambda} \alpha_B\, \mu_\Lambda[\operatorname{Ent}_B(f)]\,, \tag{2.14}$$

that is μ_Λ satisfies the general block factorization of entropy with constant $C = 1$.

The above statement is a simple consequence of the well known weighted version of Shearer inequality for the Shannon entropy; see e.g. [CHV20, Theorem 6.2] for a proof of the latter, and see e.g. [CMT15, Proposition 2.6] to go from Shannon entropy to relative entropy.

The following properties will also be used.

Lemma 2.7. *Let $\Lambda = A \cup B$ and assume that μ_Λ is a product $\mu_\Lambda = \mu_A \otimes \mu_B$. Then, for all $f \ge 0$:*

$$\operatorname{Ent}_\Lambda(\mu_B(f)) = \mu_\Lambda[\operatorname{Ent}_A(\mu_B(f))], \tag{2.15}$$

and for all $U \subset B$,

$$\mu_\Lambda[\operatorname{Ent}_A(\mu_B(f))] \le \mu_\Lambda[\operatorname{Ent}_A(\mu_U(f))]. \tag{2.16}$$

Proof. From the decomposition in Lemma 2.3 it follows that

$$\operatorname{Ent}_\Lambda(\mu_B(f)) - \mu_\Lambda[\operatorname{Ent}_A(\mu_B(f))] = \operatorname{Ent}_\Lambda(\mu_A\mu_B(f)) = \operatorname{Ent}_\Lambda(\mu_\Lambda(f)) = 0.$$

This proves (2.15). To prove (2.16) notice that by definition

$$\mu_\Lambda\left[\operatorname{Ent}_A(\mu_B(f))\right] = \mu_\Lambda\left[\mu_B(f)\log\left(\frac{\mu_B(f)}{\mu_A\mu_B(f))}\right)\right].$$

For any $U \subset B$, $\mu_B(f) = \mu_B\mu_U(f)$ and the product structure $\mu_\Lambda = \mu_A \otimes \mu_B$ implies the commutation relation $\mu_A\mu_B\mu_U = \mu_B\mu_A\mu_U$. Therefore,

$$\begin{aligned}
\mu_\Lambda\left[\operatorname{Ent}_A(\mu_B(f))\right] &= \mu_\Lambda\left[\mu_B\mu_U(f)\log\left(\frac{\mu_B\mu_U(f)}{\mu_B\mu_A\mu_U(f)}\right)\right] \\
&= \mu_\Lambda\left[\mu_U(f)\log\left(\frac{\mu_B\mu_U(f)}{\mu_B\mu_A\mu_U(f))}\right)\right] \\
&= \mu_\Lambda\left[\mu_A\left[\mu_U(f)\log\left(\frac{\mu_B\mu_U(f)}{\mu_A\mu_B\mu_U(f)}\right)\right]\right].
\end{aligned}$$

It remains to observe that

$$\mu_A\left[\mu_U(f)\log\left(\frac{\mu_B\mu_U(f)}{\mu_A\mu_B\mu_U(f)}\right)\right] \le \operatorname{Ent}_A(\mu_U(f)).$$

The latter estimate follows from the well known variational principle

$$\operatorname{Ent}_A(g) = \sup\left\{\mu_A(gh)\,,\ \mu_A(e^h) \le 1\right\} \tag{2.17}$$

valid for any A and any function $g \ge 0$; see, e.g. [Led99, Proposition 2.2]. □

2.4 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

$$P_\alpha f = \sum_{B\subset V} \alpha_B \, \mu_B(f)\,. \tag{2.18}$$

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 (2.18) is the Glauber dynamics for μ.

The α-weighted heat bath block dynamics (2.18) defines a reversible pair (P_α, μ). Moreover, its Dirichlet form satisfies

$$\mathcal{D}_\alpha(f,g) = \sum_{B\subset V} \alpha_B \, \mu[f(1-\mu_B)g] = \sum_{B\subset V} \alpha_B \, \mu\left[\mathrm{Cov}_B(f,g)\right]\,, \tag{2.19}$$

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

Lemma 2.8. *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 $\mu_* = \min_{\sigma\in\Omega}\mu(\sigma)$.*

Proof. In view of Lemma 2.7 it is sufficient to prove item 2. We note that the relative entropy decay with rate δ is equivalent to the entropy contraction

$$\mathrm{Ent}(P_\alpha f) \le (1-\delta)\,\mathrm{Ent}(f), \tag{2.20}$$

for all $f \ge 0$. By convexity of $x \mapsto x\log x$ one has

$$\begin{aligned}\mathrm{Ent}(P_\alpha f) &= \mu[P_\alpha f \log(P_\alpha f)] - \mu[f]\log\mu[f] \\ &\le \sum_B \alpha_B\,\mu[\mu_B(f)\log(\mu_B(f))] - \mu[f]\log\mu[f] = \sum_B \alpha_B\,\mathrm{Ent}(\mu_B(f)).\end{aligned} \tag{2.21}$$

From the decomposition in Lemma 2.3 it follows that

$$\mathrm{Ent}(P_\alpha f) \le \mathrm{Ent}(f) - \sum_B \alpha_B \mu[\mathrm{Ent}_B(f)]. \tag{2.22}$$

By Definition 2.5, $\sum_B \alpha_B \mu[\mathrm{Ent}_B(f)] \ge (\delta(\alpha)/C)\,\mathrm{Ent}(f)$, and therefore

$$\mathrm{Ent}(P_\alpha f) \le (1-\delta(\alpha)/C)\,\mathrm{Ent}(f). \qquad \square$$

3 Uniform block factorization implies general block factorization

We provide in this section the proofs of Lemmas 1.18 and 1.17. Recall that these are the key ingredients for proving Theorem 1.16.

Proof of Lemma 1.18. Let $\alpha = (\alpha_B)_{B\subset V}$ be a probability distribution over the subsets of V. Observe that for all $j = 1,...,k$ and all $\tau \in \Omega_{V\setminus V_j}$, $\mu_{V_j}^\tau$ is a product measure on $\Omega_{V_j}^\tau$. Therefore, we can apply Lemma 2.6 with $\Lambda = V_j$ and $\hat\alpha = (\hat\alpha_U)_{U\subset V_j}$, where $\hat\alpha_U = \omega^{-1}\sum_{B\subset V}\alpha_B \mathbf{1}(V_j\cap B = U)$ and $\omega = \sum_{B\subset V}\alpha_B\mathbf{1}(V_j\cap B\neq\emptyset)$. We get

$$\delta(\hat\alpha)\,\mathrm{Ent}_{V_j}^\tau(f) \le \sum_{U\subset V_j}\hat\alpha_U\,\mu_{V_j}^\tau[\mathrm{Ent}_U(f)] = \omega^{-1}\sum_{B\subset V}\alpha_B\,\mu_{V_j}^\tau[\mathrm{Ent}_{V_j\cap B}(f)]. \tag{3.1}$$

Observe that

$$\omega\delta(\hat{\alpha}) = \omega \min_{x\in V_j} \sum\nolimits_{U\subset V_j: U\ni x} \hat{\alpha}_U = \min_{x\in V_j} \sum\nolimits_{B\subset V: B\ni x} \alpha_B \geq \delta(\alpha),$$

and from (2.12) we have $\mu[\mathrm{Ent}_{V_j\cap B}(f)] \leq \mu[\mathrm{Ent}_B(f)]$. Hence, taking expectation in (3.1) with respect to μ we obtain

$$\delta(\alpha)\,\mu[\mathrm{Ent}_{V_j}(f)] \leq \sum_{B\subset V} \alpha_B\,\mu[\mathrm{Ent}_B(f)].$$

Summing over j we have, for all $f:\Omega\to\mathbb{R}_+$,

$$\delta(\alpha)\sum_{j=1}^{k}\mu[\mathrm{Ent}_{V_j}(f)] \leq \sum_{j=1}^{k}\sum_{B\subset V} \alpha_B\,\mu[\mathrm{Ent}_B(f)],$$

and since by assumption k-partite factorization of entropy holds with constant C, we have

$$\delta(\alpha)\mathrm{Ent}(f) \leq C\sum_{j=1}^{k}\sum_{B\subset V} \alpha_B\,\mu[\mathrm{Ent}_B(f)] \leq C\,k\sum_{B\subset V} \alpha_B\,\mu[\mathrm{Ent}_B(f)].$$

Hence, GBF holds with constant Ck. □

Proof of Lemma 1.17. Since $\lceil\theta n\rceil$-UBF holds by assumption, setting $C = C_{\mathrm{ubf}}$ one has

$$\mathrm{Ent}(f) \leq \frac{C}{\theta}\,\mathbb{E}\left[\mu\left[\mathrm{Ent}_S(f)\right]\right], \tag{3.2}$$

where S is a random set with uniform distribution over all subsets of V of cardinality $\lceil\theta n\rceil$, and $\mathbb{E}$ denotes the corresponding expectation.

Let $S_1, S_2, \ldots$ denote the connected components of S in G (taken in some arbitrary order) and for $i>1$ let $S_{<i} = \cup_{j=1}^{i-1} S_j$. Then $\mu_{S_{<i+1}}$ has the product structure $\mu_{S_{<i+1}} = \otimes_{j=1}^{i}\mu_{S_j}$. By Lemmas 2.3 and 2.7, one has the decomposition

$$\mu\left[\mathrm{Ent}_S(f)\right] = \sum_{i\geq 1}\mu\left[\mathrm{Ent}_{S_{<i+1}}(\mu_{S_{<i}}(f))\right] = \sum_{i\geq 1}\mu[\mathrm{Ent}_{S_i}(\mu_{S_{<i}}(f))], \tag{3.3}$$

where we have used Eq. (2.15) with $A = S_i$ and $B = S_{<i}$. For $\tau\in\Omega_{V\setminus S_i}$, let $\Gamma(S_i,\tau)$ be the optimal constant so that

$$\mathrm{Ent}^{\tau}_{S_i}(\mu_{S_{<i}}(f)) \leq \Gamma(S_i,\tau)\sum_{j=1}^{k}\mu^{\tau}_{S_i}\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{S_{<i}}(f))\right].$$

Let $\Gamma(S_i) = \max_{\tau\in\Omega_{V\setminus S_i}}\Gamma(S_i,\tau)$. Then,

$$\mu\left[\mathrm{Ent}_S(f)\right] \leq \sum_{i\geq 1}\Gamma(S_i)\sum_{j=1}^{k}\mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{S_{<i}}(f))\right].$$

We observe next that for all $j=1,\ldots,k$ one has

$$\mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{S_{<i}}(f))\right] \leq \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]. \tag{3.4}$$

To see this, we apply Lemma 2.7 with $A = V_j\cap S_i$, $B = S_{<i}$ and $U = V_j\cap S_{<i}$. Since $\mu_{S_{<i+1}} = \otimes_{j=1}^{i}\mu_{S_j}$ the assumptions for that lemma are satisfied and we obtain (3.4) from Eq. (2.16).

Summarizing, we have obtained

$$\mathrm{Ent}(f) \le \frac{C}{\theta} \sum_{j=1}^{k} \mathbb{E}\left[\sum_{i\ge 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right]. \tag{3.5}$$

We show next that for all $j = 1, ..., k$

$$\mathbb{E}\left[\sum_{i\ge 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right] \le C'\mu\left[\mathrm{Ent}_{V_j}(f)\right], \tag{3.6}$$

with $C' = O\left(\frac{\log(1/b)}{b^2}\Delta^2\right)$. Combined with (3.5), this concludes the proof of the lemma.

Let us fix j and let $v_1, v_2, \dots$ denote an ordering of the sites in $V_j \cap S$ such that $v_1, ..., v_{|V_j\cap S_1|}$ is an ordering of $V_j \cap S_1$, $v_{|V_j\cap S_1|+1}, ..., v_{|V_j\cap S_1|+|V_j\cap S_2|}$ is an ordering of $V_j \cap S_2$ and so on. Since, for all $i \ge 1$, $\mu_{V_j\cap S_i}$ is a product measure, Lemmas 2.3 and 2.7 (as in (3.3)) imply

$$\mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right] = \sum_{h=|V_j\cap S_1|+\cdots+|V_j\cap S_{i-1}|+1}^{|V_j\cap S_1|+\cdots+|V_j\cap S_i|} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right],$$

where ϱ_{v_h} is the conditional distribution obtained from μ by freezing the spins at all the sites outside V_j, together with all the sites $v_h, v_{h+1}, \dots, v_{|V_j\cap S|}$.

Using this decomposition and rearranging one finds

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\ge 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right] &(3.7)\\
&= \mathbb{E}\left[\sum_{i\ge 1} \Gamma(S_i) \sum_{h=|V_j\cap S_1|+\cdots+|V_j\cap S_{i-1}|+1}^{|V_j\cap S_1|+\cdots+|V_j\cap S_i|} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right]\right]\\
&= \mathbb{E}\left[\sum_{h} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right] \Gamma(S(v_h))\right], &(3.8)
\end{aligned}$$

where $S(v_h)$ denotes the (unique) connected component of S containing v_h. Notice that for each realization of S, $\mu_{V_j\cap S}$ is a product measure and so one has from Lemmas 2.3 and 2.7 that

$$\sum_{h} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right] = \mu\left[\mathrm{Ent}_{V_j\cap S}(f))\right] \le \mu\left[\mathrm{Ent}_{V_j}(f)\right];$$

the inequality follows from (2.12).

Observe that each term $\mu[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))]$, as well as the sequence $\{v_h\}$, depends on the realization S only through $V_j \cap S$. Therefore,

$$\mathbb{E}\left[\sum_{h} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right] \Gamma(S(v_h))\right] = \mathbb{E}\left[\sum_{h} \mu\left[\mathrm{Ent}_{v_h}(\varrho_{v_h}(f))\right] \mathbb{E}\left[\Gamma(S(v_h)) \,|\, V_j \cap S\right]\right],$$

where $\mathbb{E}\left[\Gamma(S(v_h)) \,|\, V_j \cap S\right]$ is the conditional expectation of $\Gamma(S(v_h))$ given the realization $V_j \cap S$. Therefore, (3.6) follows if we prove that

$$\max_{W\subset V_j} \max_{v\in W} \mathbb{E}\left[\Gamma(S(v)) \,|\, V_j \cap S = W\right] \le C'. \tag{3.9}$$

Now, for a b marginally bounded spin system, it follows from Lemma 4.2 in [CLV21] and (2.12) that

$$\Gamma(S(v)) \le \zeta |S(v)|^3 z^{|S(v)|},$$

where $\zeta = \zeta(b) = \frac{3\log(1/b)}{2b^2}$ and $z = 1/b^2$. Thus,

$$\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[\Gamma(S(v))\,|\,V_j\cap S=W\right] \le \zeta\cdot\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[|S(v)|^3 z^{|S(v)|}\;\middle|\;V_j\cap S=W\right]. \tag{3.10}$$

To bound the expectation on the right-hand-side of (3.10), we consider the graph G_2 with vertex set V and edge set $E\cup E_2$, where E is the edge set of G and E_2 is the set of all pairs of vertices with a common neighbor in G. Note that G_2 has maximum degree Δ^2. Let $\mathcal{A}_v(a)$ be the collection of subsets of vertices $U\subset V$ such that $|U|\ge a$, $v\in U$ and the induced subgraph $G_2[U]$ of U in G_2 is connected.

Now, let us fix the set $W = V_j\cap S$ and the vertex $v\in W$ and let $S_2 := (S(v)\cap V_{\neq j})\subset S$, where $V_{\neq j} := \bigcup_{i:i\neq j} V_i$. We claim that when the event $\{|S(v)| = a\}$ occurs for some $a\in\mathbb{N}$, then $S_2\in\mathcal{A}_v(\frac{a}{\Delta+1})$. Indeed, $G_2[S_2]$ is connected, since $S(v)$ is connected in G and removing the vertices in V_j from $S(v)$ will not disconnect S_2 in G_2. Moreover, since each vertex in $S(v)\cap V_j$ is a neighbor of some vertex in $S(v)\cap V_{\neq j}$ and the maximum degree of G is Δ, one has $\Delta|S(v)\cap V_{\neq j}|\ge|S(v)\cap V_j|$, and so

$$a = |S(v)\cap V_j| + |S(v)\cap V_{\neq j}| \le (\Delta+1)|S(v)\cap V_{\neq j}|,$$

which implies that $|S_2| = |S(v)\cap V_{\neq j}|\ge a/(\Delta+1)$. Given S, let $T_2(v)$ denote the connected component of S in G_2 containing v, and note that $S_2\subset T_2(v)$. Then, for any $W\subset V_j$, $v\in W$ and integer $a\ge 1$ we get

$$\begin{aligned}\mathbb{P}\left(|S(v)| = a\,|\,V_j\cap S = W\right) &\le \mathbb{P}\left(\exists\, S'\in\mathcal{A}_v\left(\frac{a}{\Delta+1}\right); S'\subset S\right)\\ &\le \mathbb{P}\left(|T_2(v)|\ge\frac{a}{\Delta+1}\right).\end{aligned} \tag{3.11}$$

To estimate the size of the connected component $T_2(v)$ one can use Lemma 4.3 from [CLV21], which implies that for any integer $m\ge 1$,

$$\mathbb{P}\left(|T_2(v)| = m\right)\le\frac{\ell}{n}(2e\Delta^2\theta)^{m-1}. \tag{3.12}$$

Indeed, the only difference with respect to Lemma 4.3 from [CLV21] is that we have maximum degree Δ^2 here instead of Δ. In particular, if $2e\Delta^2\theta\le 1/2$, using $\frac{\ell}{n}\le 2\theta$,

$$\mathbb{P}\left(|T_2(v)|\ge\frac{a}{\Delta+1}\right)\le 4\theta(2e\Delta^2\theta)^{\lfloor\frac{a}{\Delta+1}\rfloor-1}\le\Delta^{-2}(2e\Delta^2\theta)^{\lfloor\frac{a}{\Delta+1}\rfloor}. \tag{3.13}$$

It follows that

$$\mathbb{E}\left[|S(v)|^3 z^{|S(v)|}\,|\,V_j\cap S=W\right] = \sum_{a\ge1} a^3 z^a\cdot\mathbb{P}\left(|S(v)| = a\,|\,V_j\cap S=W\right) \tag{3.14}$$

$$\le\Delta^{-2}\sum_{a\ge1}a^3(2e\Delta^2\theta z^{\Delta+1})^{\lfloor\frac{a}{\Delta+1}\rfloor}\le C_1\Delta^2, \tag{3.15}$$

for some absolute constant C_1 provided that $2e\Delta^2\theta z^{\Delta+1}\le 1/2$. The last bound can be seen e.g. by writing the sum over a as a sum over ℓ and the by summing over $(\ell-1)(\Delta+1)\le a\le\ell(\Delta+1)-1$. This implies that

$$\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[|S(v)|^3 z^{|S(v)|}\;\middle|\;V_j\cap S=W\right]\le C_1\Delta^2.$$

Hence, (3.9) and (3.6) hold with $C' = C_1\zeta\Delta^2$, and consequently k-partite factorization holds with constant $C_{\text{ubf}}C_1\zeta\Delta^2/\theta$. □

4 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 1.10 concerned with Glauber dynamics and Hamming metric in Section 4.1; this will serve as a concrete example to illustrate our approach for establishing spectral independence. In Section 4.2, we consider arbitrary metric and prove Theorem 1.10. Finally, we consider general Markov chains and metrics in Section 4.3 and prove Theorem 1.11.

4.1 Warm-up: contraction for Glauber dynamics and Hamming metric

In this section, we prove a simpler version of Theorem 1.10, 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 4.1. *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 4.2. 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 4.1 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 4.2.1, the contraction rate satisfies $\kappa_\ell \le 1 - \epsilon/\ell$, so we eventually get $\eta \le 2/\epsilon$ just as from Theorem 4.1.

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 [ALO20, CLV20, CGvV21, FGYZ21, CLV21] is to upper bound the infinity norm of J. More specifically, for each $(x,a) \in \mathcal{X}$ we define

$$S(x,a) = \sum_{(y,a') \in \mathcal{X}} |J(x,a;y,a')| \tag{4.1}$$

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 prove a suitable upper bound on $S(x,a)$. 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

$$f(\sigma) = \sum_{(y,a')\in\mathcal{X}} t(x,a;y,a')\,\mathbf{1}_{\{\sigma_y = a'\}}, \tag{4.2}$$

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 in the following functional form, which follows from Kantorovich-Rubinstein duality [Vil21],

$$W_{1,d}(\mu,\nu) = \sup_{\substack{f:\Omega\to\mathbb{R} \\ L_d(f)\le 1}} \mathbb{E}_\mu f - \mathbb{E}_\nu f. \tag{4.3}$$

Observe that, the function f defined by (4.2) 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 (4.3) 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 $W_{1,d_{\mathrm{H}}}(\nu,\mu) \le C$ for μ and $\nu = \mu(\cdot \mid \sigma_x = a)$ for any pair (x,a), then $\lambda_1(J) \le 2C$ and the η-spectral independence with $\eta = 2C$ would follow.

The following lemma, which generalizes previous works [BN19, RR19], 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 [BN19, Theorem 3.1] and [RR19, 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 4.3.1.

Lemma 4.3. *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 4.3 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 4.3. The proof imitates the arguments from [BN19, RR19]. 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,

$$h = \sum_{t=0}^{\infty} P^t \bar{f}. \tag{4.4}$$

See Lemma 2.1 in [BN19] 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 (4.3) 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 (4.3). Since μ is κ-contractive w.r.t. P and d, for all $\sigma, \tau \in \Omega$ and every integer $t \geq 1$ we have

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

We then deduce that

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

This implies that $L_d(h) \leq L_d(f)/(1-\kappa)$ 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 (4.4) is well-defined; i.e., the series on the right-hand side of (4.4) is convergent. The rest of the proof does not require the irreducibility of P. In fact, one can deduce the convergence of (4.4) solely from the contraction of P. Note that for all $\sigma \in \Omega$,

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

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 (4.4) 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

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

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 4.3, we can now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 4.3 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \leq \frac{L_{d_{\mathrm{H}}}(f)}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot), Q(\sigma,\cdot))\right] \tag{4.5}$$

where $S(x,a)$ is given by (4.1), f is given by (4.2), 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 $\sigma \in \Omega^{(x,a)}$,

$$W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot), Q(\sigma,\cdot)) \leq \frac{1}{n}. \tag{4.6}$$

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 $d_{\mathrm{H}}(\sigma_1,\sigma_2) \leq 1$ 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 $L_{d_{\mathrm{H}}}(f) \leq 2$ and (4.6), we obtain from (4.5) that $S(x,a) \leq \frac{2}{(1-\kappa)n}$ for each (x,a); consequently, $\lambda_1(J) \leq \frac{2}{(1-\kappa)n}$. The same argument holds for μ^τ under any pinning τ as well, and spectral independence then follows. □

4.2 Contraction for Glauber dynamics and general metrics

In this section, we generalize the Hamming metric assumption in Theorem 4.1 to any weighted Hamming metric or any metric equivalent to Hamming, which establishes Theorem 1.10. We restate it here for convenience.

Theorem 1.10. *(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 1.10 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

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

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

Unfortunately, the proof of Theorem 4.1 does not work directly in this scenario. The reason is that the right-hand side of (4.5), 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 1.10(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

$$S_w(x,a) = \sum_{(y,a') \in \mathcal{X}} w(y)\, |J(x,a;y,a')|. \tag{4.7}$$

Such weighted sums were considered in [CLV20, 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

$$f_w(\sigma) = \sum_{(y,a') \in \mathcal{X}} w(y)\, t(x,a;y,a')\, \mathbf{1}_{\{\sigma_y = a'\}} \tag{4.8}$$

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 4.3 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 1.10. 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 4.1, and in particular the right-hand side of (4.9) below (analogous to (4.5)) can be upper bounded using the γ-equivalence.

Proof of Theorem 1.10(2). For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 4.3 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le \frac{L_d(f)}{1-\kappa} \mathbb{E}_\nu \left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \right] \tag{4.9}$$

where $S(x,a)$ and f are defined by (4.1), (4.2) 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 (4.6). Thus, we obtain from (4.9) that $S(x,a) \le \frac{2\gamma^2}{(1-\kappa)n}$. The rest of the proof is the same as Theorem 4.1. □

4.2.1 Application: Dobrushin uniqueness condition

As an application of Theorem 1.10, we show that the Dobrushin uniqueness condition, as well as its generalizations [Hay06, DGJ09], implies spectral independence. Recall that the Dobrushin dependency matrix R is a $|V| \times |V|$ matrix defined as $R(x,x) = 0$ and

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

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 1.13 from the introduction.

Theorem 1.13. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \le 1 - \epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

Remark 4.4. If $\|R\|_\infty < 1$, then the Glauber dynamics mixes rapidly by a simple application of the path coupling method of Bubley and Dyer [BD97]. The same is true under the Dobrushin uniqueness condition, i.e., when $\|R\|_1 < 1$. Hayes [Hay06] generalized the condition to the spectral norm $\|R\|_2 < 1$. Dyer, Goldberg, and Jerrum [DGJ09] 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 1.13 is technically better than previous works since for a nonnegative matrix R one has $\varrho(R) \le \|R\|$ for any matrix norm, and the inequality can be strict for all norms when R is not irreducible; see [DGJ09] 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 4.5 ([DGJ09, Lemma 20])**.** *If $w \in \mathbb{R}_+^V$ is a positive vector such that $Rw \le (1-\epsilon)w$ 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 4.6 ([Mey00, Example 7.10.2])**.** *If $M, N \in \mathbb{R}_+^{n\times n}$ are two nonnegative square matrices such that $M \le N$ entrywisely, then $\varrho(M) \le \varrho(N)$.*

We give below the proof of Theorem 1.13.

Proof of Theorem 1.13. 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 $Rw = \varrho(R)w \le (1-\epsilon)w$ and $w > 0$ by the Perron-Frobenius theorem. Hence, Lemma 4.5 and (the proof of) Theorem 1.10(1) immediately yield $\lambda_1(J) \le 2/\epsilon$. 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, $Rw_\delta \le R_\delta w_\delta = \varrho(R_\delta)w_\delta$. 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 [Ale13]. Thus, $\varrho(R_\delta) < 1$ for sufficiently small δ. Then by Lemma 4.5 and Theorem 1.10(1), for δ small enough, we have $\lambda_1(J) \le 2/(1-\varrho(R_\delta))$. Taking $\delta \to 0$ and using the assumption that $\varrho(R) \le 1-\epsilon$, we obtain $\lambda_1(J) \le 2/\epsilon$.

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 $R^\tau(x,y) \le R(x,y)$ for all $x, y \in V$. We deduce from Lemma 4.6 that $\varrho(R^\tau) \le \varrho(R) \le 1-\epsilon$ and thus this is reduced to the no-pinning case. Therefore, we get $\lambda_1(J^\tau) \le 2/\epsilon$ for all τ and spectral independence then follows. □

4.3 Contraction for general Markov chains and general metrics

In this section, we generalize Theorem 4.1 to arbitrary "local" Markov chains and arbitrary metrics close to the Hamming metric. In particular, we prove Theorem 1.11.

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 4.7. *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 4.8. *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 4.1 and 1.10(2). For an arbitrary pinning τ and $(x,a) \in \mathcal{X}^{\tau}$, we define

$$S^{\tau}(x,a) = \sum_{(y,a') \in \mathcal{X}^{\tau}} |J^{\tau}(x,a;y,a')| \tag{4.10}$$

and

$$f^{\tau}(\sigma) = \sum_{(y,a') \in \mathcal{X}^{\tau}} t^{\tau}(x,a;y,a')\, \mathbf{1}_{\{\sigma_y = a'\}} \tag{4.11}$$

where $t^{\tau}(x,a;y,a') = \operatorname{sgn}(J^{\tau}(x,a;y,a'))$; these definitions are analogous to (4.1) and (4.2) with pinning τ. Let $\tau' = \tau \cup (x,a)$. Then we deduce from Lemma 4.3 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 1.10(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 4.9. 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 4.2 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

$$D = \max_{p_t} \max_{x} \sum_{B \in \mathcal{B}: x \in B} p_t(B), \tag{4.12}$$

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 4.8 we establish Theorem 1.11, which we restate here for convenience.

Theorem 1.11. *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 4.8. 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 $d_{\mathrm{H}}(\sigma_1, \sigma_2) \leq |B| \leq M$. Therefore,

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

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

Remark 4.10. 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 1.11 the maximum block size M can be replaced by the maximum component size of all blocks.

4.3.1 Application: flip dynamics for colorings

In this section we establish spectral independence for colorings utilizing Theorem 1.11.

Theorem 4.11. *Let $\epsilon_0 \approx 10^{-5} > 0$ be a fixed constant. Let $\Delta, q \geq 3$ 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 1.11, we need a contractive Markov chain for sampling colorings of a graph. Vigoda considered the *flip dynamics* [Vig00] and showed that it is contractive for the Hamming metric when the number of colors $q > \frac{11}{6}\Delta$. Recently, [CDM+19] 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 [Vig00] and the recent improvement [CDM$^+$19], 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 [CDM$^+$19], where the authors established contraction of the flip dynamics using the path coupling method.

Lemma 4.12 ([CDM$^+$19])**.** *Under the assumptions of Theorem 4.11, 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 [CDM$^+$19] 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 4.2 addressing the same issue for the Glauber dynamics and also Remark 4.9 for general select-update dynamics.

We give below the proof of Theorem 4.11.

Proof of Theorem 4.11. 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 $\operatorname{dist}(x,y) \leq 5$, which happens with probability at most Δ^6/n. The theorem then follows from Lemma 4.12 and Theorem 1.11. □

We conclude here with the proof of Theorem 1.1.

Proof of Theorem 1.1. By Theorem 4.11 the uniform distribution μ of proper colorings is spectrally independent. Then the results follows immediately from Theorem 1.7. □

4.3.2 Application: block dynamics for Potts model

Here we apply Theorems 4.1 and 1.11 to the ferromagnetic Potts model to establish spectral independence.

Theorem 4.13. *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 [Ull14] and [BGP16] regarding the contraction of the Glauber dynamics and of the heat-bath block dynamics with a specific choice of blocks.

Lemma 4.14 ([Ull14, Corollary 2.14] & [BGP16, Proposition 2.2])**.** *Under the assumptions in Part 1 of Theorem 4.13, 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 4.15 ([BGP16, Theorem 2.7])**.** *Under the assumptions in Part 2 of Theorem 4.13, 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 4.16. To be more precise, [BGP16] shows that the conclusion of Lemma 4.15 is true when β, q, and the maximum block size $M = \max_{x\in V} |B_x|$ satisfy

$$\beta\left(\Delta - 1 + \frac{1}{M}\right) + 3M(\ln\Delta + \ln M) \le \ln q. \tag{4.13}$$

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 4.13 implies (4.13). 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 4.13 is an immediate consequence of Lemmas 4.14, 4.15 and the results proved in this section.

Proof of Theorem 4.13. Part 1 follows directly from Lemma 4.14 and Theorem 4.1. For Part 2, we note that the block dynamics from Lemma 4.15 corresponds to a select-update dynamics with $M = O(1/\delta)$ and $D = \Delta^{O(1/\delta)}/n$; to see the bound on D, we observe that if $x \in B_y$ for some y then the graph distance between x and y is at most $|B_y| = O(1/\delta)$ since $G[B_y]$ is connected, and hence the number of blocks containing x is at most $\Delta^{O(1/\delta)}$. The theorem then follows from Lemma 4.15 and Theorem 1.11. □

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 [CLV20]. For Potts model, Theorem 4.13 establishes spectral independence in the corresponding parameter regimes. The theorem then follows from Theorems 1.7 and 1.8. □

5 Spectral independence and entropy factorization

The goal of this section is to reformulate in the setting of spin systems some of the key facts that were derived in [CLV21] and the references therein in the framework of simplicial complexes. This specialization yields some minor simplification in the main proofs, and may be of use for later reference. The approach consists in exploiting a recursive scheme which allows one to derive a global contraction estimate by analysing the spectral norm of a local operator. This is reminiscent of the recursive approach developed in [CCL03, CM03, Cap04], where similar ideas were used to derive spectral gap estimates for a class of conservative spin systems. The argument here is more robust and, unlike the one in [CCL03, CM03, Cap04], it does not rely on symmetries of the underlying measures.

We first introduce some notation. Let f be a function of the spin configuration σ in the whole region V, and $U \subset V = [n]$ a subset of vertices. Recall the notation $\mu_{V\setminus U}$ for the conditional distribution given the spins in U, and write $\mathrm{Av}_{|U|=\ell}$ for the uniform average over all sets $U \subset [n]$ with ℓ vertices. We are going to prove the following result that was established in [CLV21].

Theorem 5.1. *If the spin system is η-spectrally independent and b-marginally bounded then there exists a constant $C = O(1 + \frac{\eta}{b})$ such that for any $\ell = \{1, \dots, n-1\}$ and for all $f \geq 0$:*

$$\frac{n}{\ell} \operatorname{Av}_{|U|=\ell} \operatorname{Ent}(\mu_{V\setminus U} f) \leq C \operatorname{Ent} f. \tag{5.1}$$

Moreover, for any $\theta \in (0, 1]$, there exists $C = \left(\frac{1}{\theta}\right)^{O(\frac{\eta}{b})}$ such that for $\ell = \lceil \theta n \rceil$:

$$\frac{\ell}{n} \operatorname{Ent} f \leq C \operatorname{Av}_{|\Lambda|=\ell} \mu \left[\operatorname{Ent}_\Lambda f\right]. \tag{5.2}$$

We remark that, when $\ell = 1$, (5.1) takes the form of an approximate subadditivity statement:

$$\sum_{x \in V} \operatorname{Ent}(f_x) \leq C \operatorname{Ent}(f), \tag{5.3}$$

with constant $C = O(1+\frac{\eta}{b})$. Here the functions f_x are defined by $f_x = \mu(f \,|\, \sigma_x) = \mu_{V\setminus\{x\}} f$. When $\mu(f) = 1$ then $\nu = f\mu$ is a probability measure and, if μ_x denotes the marginal of μ on x, then $f_x \mu_x$ gives the marginal of ν on x. The inequality (5.3) is known to be equivalent to a Brascamp-Lieb type inequality for the measure μ; see [CLL04, CC09]. For a general discussion of subadditivity of entropy, Brascamp-Lieb type inequalities, and their applications, see for instance [BCLM11] and the references therein. On the other hand (5.2) is the uniform block factorization statement ℓ-UBF with $\ell = \lceil \theta n \rceil$; see Definition 1.15.

We articulate the proof in two steps. The first is a recursive scheme which allows one to go from a local inequality to a global one; see Lemma 5.3. The second step is a control of the local inequality; see Lemma 5.4.

5.1 Setting up the recursion

If $U \subset V$, and $\tau = \tau_U$ a configuration of spins on U, recall that we use notation $\mu^\tau = \mu(\cdot \,|\, \tau)$ for the conditional distribution $\mu_{V\setminus U}$ when the spins on U are given by τ. Moreover, we write $\mu^{\tau,x} = \mu(\cdot \,|\, \tau \cup \sigma_x)$ if we additionally condition on the spin σ_x at vertex $x \notin U$ and similarly for $\mu^{\tau,x,y} = \mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$ for $x, y \notin U$, so that e.g. the expression $\mu^\tau \left[\operatorname{Ent}_{\mu^{\tau,x,y}} f\right]$ indicates the entropy of f with respect to $\mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$,

$$\operatorname{Ent}_{\mu^{\tau,x,y}} f = \mu^{\tau,x,y}[f \log(f/\mu^{\tau,x,y}(f))]$$

averaged over the two spins σ_x, σ_y sampled according to μ^τ. Define the constants α_k, $k = 0, \dots, n-2$, as the largest numbers such that the inequalities

$$(1 + \alpha_k) \operatorname{Av}_{x \notin U} \operatorname{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \leq \operatorname{Av}_{x,y \notin U} \operatorname{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)), \tag{5.4}$$

hold for all $k = 0, \dots, n-2$, for all $U \subset [n]$ with $|U| = k$, for all configurations τ on U and for all functions $f \geq 0$. The symbol $\operatorname{Av}_{x \notin U}$ denotes the uniform average over all $n - k$ vertices $x \notin U$, and $\operatorname{Av}_{x,y \notin U}$ stands for the uniform average over all $(n-k)(n-k-1)$ pairs (x, y) with $x, y \notin U$ and $x \neq y$. We refer to (5.4) as the local inequality, since for each choice of x, y, the distributions involved are concerned with the spins at two vertices only.

Remark 5.2. Fix $x, y \notin U$. Using $\mu^{\tau,x} f = \mu^{\tau,x} \mu^{\tau,x,y} f$, from Lemma 2.3 we have the decomposition

$$\operatorname{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) = \operatorname{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mu^\tau \left[\operatorname{Ent}_{\mu^{\tau,x}}(\mu^{\tau,x,y}(f)\right].$$

In particular, $\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \geq \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))$ and therefore (5.4) is always true with $\alpha_k = 0$. If μ is a product measure then the subadditivity of entropy for product measures gives

$$\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \geq \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,y}(f)),$$

which implies the validity of (5.4) with $\alpha_k = 1$ for all $k = 0, \dots, n-2$.

The recursion is based on the following statement, which rephrases [CLV21, Theorem 5.4].

Lemma 5.3. *Let α_k, $k = 0, \dots, n-2$, be defined by* (5.4)*. Then, for all functions $f \geq 0$,*

$$\mathrm{Av}_{|U|=j}\mathrm{Ent}(\mu_{V\setminus U} f) \leq (1-\kappa_j)\mathrm{Ent}(f), \qquad j = 1, \dots, n-1, \tag{5.5}$$

where

$$\kappa_j = \frac{\sum_{i=j}^{n-1}\Gamma_i}{\sum_{i=0}^{n-1}\Gamma_i}, \qquad \Gamma_i = \prod_{k=0}^{i-1}\alpha_k, \quad \Gamma_0 = 1.$$

Proof. The claim (5.5) follows from the fact that for all $k = 1, \dots, n-1$:

$$\mathrm{Av}_{|U|=k}\,\mathrm{Ent}(\mu_{V\setminus U} f) \leq \delta_k \mathrm{Av}_{|U|=k+1}\,\mathrm{Ent}(\mu_{V\setminus U} f), \qquad \delta_k = \frac{\sum_{i=0}^{k-1}\Gamma_i}{\sum_{i=0}^{k}\Gamma_i}, \tag{5.6}$$

since $\mathrm{Av}_{|U|=n}\,\mathrm{Ent}(\mu_{V\setminus U} f) = \mathrm{Ent}(f)$, and $\delta_j\delta_{j+1}\cdots\delta_{n-1} = (1-\kappa_j)$.

To prove (5.6), note that it holds for $k=1$ with $\delta_1 = 1/(1+\alpha_0) = \Gamma_0/(\Gamma_0+\Gamma_1)$ by the assumption (5.4) at $\tau = \emptyset$. Next, we suppose it holds for $0 < k-1 < n-1$ and show it for k. For any $|U| = k+1$ and $U' \subset U$ with $|U'| = k-1$, setting $\{x,y\} = U \setminus U'$ and letting $\tau = \tau_{U'}$ be the configuration on U', as in Lemma 2.3 we have the decomposition

$$\begin{aligned}\mathrm{Ent}(\mu_{V\setminus U} f) &= \mathrm{Ent}(\mu(\mu_{V\setminus U} f \,|\, \tau_{U'})) + \mu\left[\mathrm{Ent}(\mu_{V\setminus U} f \,|\, \tau_{U'})\right] \\ &= \mathrm{Ent}(\mu_{V\setminus U'} f) + \mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y} f)\right].\end{aligned} \tag{5.7}$$

Averaging we obtain

$$\begin{aligned}\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U} f) = \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'} f) \\ + \mathrm{Av}_{|U'|=k-1}\mathrm{Av}_{x,y\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y} f)\right].\end{aligned} \tag{5.8}$$

In the same way

$$\begin{aligned}\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U} f) = \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'} f) \\ + \mathrm{Av}_{|U'|=k-1}\mathrm{Av}_{x\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x} f)\right].\end{aligned} \tag{5.9}$$

From (5.4),

$$\begin{aligned}&\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U} f) - \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'} f) \\ &\qquad \geq (1+\alpha_{k-1})\mathrm{Av}_{|U'|=k-1}\mathrm{Av}_{x\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x} f)\right] \\ &\qquad = (1+\alpha_{k-1})\left[\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U} f) - \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'} f)\right].\end{aligned} \tag{5.10}$$

Therefore,

$$\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U} f) \geq (1+\alpha_{k-1})\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U} f) - \alpha_{k-1}\mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'} f).$$

By the inductive assumption (5.6) at $k-1$ we have

$$\begin{aligned}\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U} f) &\geq (1+\alpha_{k-1}-\alpha_{k-1}\delta_{k-1})\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U} f) \\ &= \delta_k^{-1}\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U} f).\end{aligned}$$

□

5.2 Estimating the local coefficients

The next step is an estimate on the coefficients α_k appearing in (5.4).

Lemma 5.4. *If the spin system is η-spectrally independent and b-marginally bounded then the local inequality* (5.4) *holds with*

$$\alpha_k \geq 1 - \frac{2\eta}{b(n-k-1)}. \tag{5.11}$$

Proof. Fix $U \subset V$, $|U| = k \leq n-2$ and $\tau = \tau_U$. We may assume $\mu^\tau(f) = 1$, which implies $\mu^\tau(\mu^{\tau,x,y}(f)) = \mu^\tau(\mu^{\tau,x}(f)) = 1$ for all $x, y \notin U$. For simplicity, we write $\mathrm{Av}_{x,y}$ and Av_x for the averages $\mathrm{Av}_{x,y\notin U}$ and $\mathrm{Av}_{x\notin U}$. Observe that

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \\
&\quad= \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x,y}(f)\log\mu^{\tau,x,y}(f) - \mu^{\tau,x}(f)\log\mu^{\tau,x}(f) - \mu^{\tau,y}(f)\log\mu^{\tau,y}(f)\right] \qquad (5.12)\\
&\quad= \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x,y}(f)\log\frac{\mu^{\tau,x,y}(f)}{\mu^{\tau,x}(f)\mu^{\tau,y}(f)}\right]. \qquad (5.13)
\end{aligned}$$

Using $a\log(a/b) \geq a - b$ for all $a, b \geq 0$,

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \\
&\quad\geq 1 - \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x}(f)\mu^{\tau,y}(f)\right] \\
&\quad= -\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right]. \qquad (5.14)
\end{aligned}$$

We may rewrite

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right] \\
&\qquad= \frac{1}{n-k-1}\sum_{(x,a)\in\mathcal{X}} \nu(x,a)\varphi(x,a)[J^\tau\varphi](x,a), \qquad (5.15)
\end{aligned}$$

where

$$\varphi(x,a) = \mu^\tau(f\,|\,\sigma_x = a) - 1 = [\mu^{\tau,x}(f)](a) - 1,$$

$\mathcal{X}$ is the set of all pairs (x,a) where $x \in V\setminus U$ (if U is the set where $\tau = \tau_U$ is specified) and $a \in [q]$, ν denotes the probability measure on $\mathcal{X}$ obtained by setting

$$\nu(x,a) = \frac{1}{n-k}\,\mu^\tau(\sigma_x = a),$$

and $J^\tau : \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$ denotes the influence matrix from Definition 1.3. Note that in the derivation of (5.15) we have used the fact that for each fixed $y \notin U$ one has

$$\sum_{a'\in[q]} \nu(y,a')\varphi(y,a') = \frac{1}{n-k}\,\mu^\tau(\mu^{\tau,y}(f)-1) = 0.$$

Observe that J^τ is self-adjoint in $L^2(\mathcal{X},\nu)$:

$$\nu(x,a)J^\tau(x,a;y,a') = \nu(y,a')J^\tau(y,a';x,a). \tag{5.16}$$

In particular, its eigenvalues are real. Let $\eta \geq 0$ denote its largest eigenvalue (the eigenvalue zero always exists since all row sums of J^τ vanish). Letting $\langle\cdot,\cdot\rangle$ denote the scalar product in $L^2(\mathcal{X},\nu)$ we have $\langle\psi, J^\tau\psi\rangle \leq \eta\langle\psi,\psi\rangle$ for all $\psi \in L^2(\mathcal{X},\nu)$. Therefore,

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right] \\
&\quad= \frac{1}{n-k-1}\langle\varphi, J^\tau\varphi\rangle \leq \frac{\eta}{n-k-1}\langle\varphi,\varphi\rangle \\
&\quad= \frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)^2\right] = \frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)). \qquad (5.17)
\end{aligned}$$

Recalling (5.14) we have shown

$$\begin{aligned}\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))\\ \geq -\frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).\end{aligned} \tag{5.18}$$

Next, observe that for every fixed $x \notin U$, setting $h^\tau(\sigma_x) = [\mu^{\tau,x}(f)](\sigma_x)$:

$$\begin{aligned}\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)) &= \sum_a \mu^\tau(\sigma_x = a)(h^\tau(a)-1)^2\\ &\leq \frac{1}{b}\left(\sum_a \mu^\tau(\sigma_x = a)|h^\tau(a)-1|\right)^2\end{aligned}$$

where $b = \min_{x\notin U}\min_a \mu^\tau(\sigma_x = a)$, as in Definition 1.5, with the minimum over a restricted to spin values that are allowed at x, that is such that $\mu^\tau(\sigma_x = a) > 0$, and we have used $\sum_i a_i^2 \leq (\sum_i a_i)^2$ for all $a_i \geq 0$. Pinsker's inequality shows that

$$\sum_a \mu^\tau(\sigma_x = a)|h^\tau(a)-1| \leq \sqrt{2\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))}.$$

It follows that

$$\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)) \leq \frac{2}{b}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)). \tag{5.19}$$

Inserting (5.19) into (5.18) concludes the proof. □

5.3 Proof of Theorem 5.1

From Lemma 5.3, we see that (5.1) holds with $C = \frac{n}{\ell}(1-\kappa_\ell)$. From Lemma 5.4 if follows that

$$\alpha_k \geq \max\{1 - R/(n-k-1), 0\}, \qquad R = \lceil 2\eta/b\rceil.$$

Using this bound in the definition of the coefficients κ_ℓ and rearranging, see Section 2.2 of [CLV21], it is not hard to see that for any $1 \leq \ell \leq n-1$:

$$\kappa_\ell \geq \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}. \tag{5.20}$$

In particular,

$$\frac{n}{\ell}(1-\kappa_\ell) \leq \frac{n}{\ell}\left(1 - \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}\right).$$

Remarkably, the expression in the right hand side above is decreasing with ℓ, and therefore it is always less than $R+1$, its value at $\ell = 1$. This shows that (5.1) holds with $C \leq R+1 = O(1+\frac{\eta}{b})$.

To prove (5.2), we start with the decomposition

$$\mathrm{Av}_{|\Lambda|=\ell}\,\mu\left[\mathrm{Ent}_\Lambda f\right] = \mathrm{Ent}(f) - \mathrm{Av}_{|U|=n-\ell}\,\mathrm{Ent}\left[\mu_{V\setminus U} f\right],$$

which follows from Lemma 2.3. Therefore Lemma 5.3 implies that (5.2) holds with $C = \frac{\ell}{n\,\kappa_{n-\ell}}$. Using (5.20) we see that

$$\frac{\ell}{n\,\kappa_{n-\ell}} \leq \frac{(n-1)\cdots(n-R)}{(\ell-1)\cdots(\ell-R)}.$$

In particular, if $\ell = \lceil \theta n\rceil$ with $\theta \in (0,1]$ fixed, then for all sufficiently large n one has $\frac{\ell}{n\,\kappa_{n-\ell}} \leq (\frac{1}{\theta})^{O(R)}$. This ends the proof of Theorem 5.1.

6 Optimal mixing of the SW dynamics

In this section, we show that for ferromagnetic Potts models, the k-partite factorization of entropy, as defined in (1.7), implies optimal mixing of the Swendsen-Wang (SW) dynamics. Since we have already established that, for any spin system, k-partite factorization is implied by spectral independence, we then deduce Theorem 1.8 from the introduction.

We again take $G = (V, E)$ to be an n-vertex graph of maximum degree Δ and μ to be the Potts distribution on G with configuration space $\Omega = [q]^V$. The SW dynamics takes a spin configuration, transforms it into a "joint" spin-edge configuration, performs a step in the joint space, and then drops the edges to obtain a new Potts configuration. Formally, from a Potts configuration $\sigma_t \in [q]^V$, a transition $\sigma_t \to \sigma_{t+1}$ of the SW dynamics is defined as follows:

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

It will be useful for us to consider the "joint" Edwards-Sokal distribution for G with parameters $p \in [0, 1]$ and integer $q \geq 2$. Let $\Omega_{\mathrm{j}} = \Omega \times \{0, 1\}^E$ be the set of "joint" spin-edge configurations (σ, A) consisting of a spin assignment to the vertices $\sigma \in \Omega$ and a subset of edges $A \subset E$. The Edwards-Sokal measure assigns to each $(\sigma, A) \in \Omega_{\mathrm{j}}$ a probability given by

$$\nu(\sigma, A) = \frac{1}{Z_{\mathrm{j}}}\, p^{|A|}(1-p)^{|E|-|A|}\mathbf{1}(\sigma \sim A), \tag{6.1}$$

where $\sigma \sim A$ means that $A \subset M(\sigma)$ (i.e., every edge in A is monochromatic in σ) and Z_{j} is the corresponding normalizing constant or partition function. When $p = 1 - e^{-\beta}$, the "spin marginal" of ν is precisely the Potts distribution μ and $Z_G = Z_{\mathrm{j}}$, and the "edge marginal" of ν corresponds to the random-cluster measure; see, e.g., [FK72, Gri06] for extensive background on these measures.

Before stating our results, we stipulate some notation. We write

$$\mathrm{Ent}_\nu(f) = \nu[f \log(f/\nu(f))]$$

for the entropy of the function $f : \Omega_{\mathrm{j}} \to \mathbb{R}_+$ with respect to ν. For a fixed configuration $\sigma \in \Omega$ and subset of edges $A \subset E$, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ denote the entropy of f with respect to the conditional measures $\nu(\cdot \,|\, \sigma)$ and $\nu(\cdot \,|\, A)$, respectively. More precisely, for a given $\sigma \in \Omega$, $\nu(\cdot \,|\, \sigma)$ is the measure ν conditioned on the event that the spin configuration is equal to σ, and for a given $A \subset E$, $\nu(\cdot \,|\, A)$ is the measure ν conditioned on the event that the edge configuration is equal to A. In this way, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ are functions of σ and A, respectively, and $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right]$, $\nu\left[\mathrm{Ent}_\nu(f \,|\, A)\right]$ denote the corresponding expectations with respect to ν. The main result in this section is stated as follows.

Theorem 6.1. *Suppose μ satisfies the k-partite factorization of entropy with constant C_{par}; see Eq.* (1.7)*. Then, there exists a constant $C = C(C_{\mathrm{par}}, \beta, \Delta)$ such that for all $f : \Omega_{\mathrm{j}} \mapsto \mathbb{R}_+$*

$$\mathrm{Ent}_\nu(f) \leq C\, \left(\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right] + \nu[\mathrm{Ent}_\nu(f \,|\, A)]\right). \tag{6.2}$$

The constant C satisfies $C = C_{\mathrm{par}} \times O(\beta\Delta^2 e^{\beta\Delta})$.

We call (6.2) the *spin/edge factorization of entropy* with constant C for the joint measure ν. The main motivation for this inequality is the result established in [BCP$^+$21, Lemma 1.8] that on any n-vertex graph, spin/edge factorization with constant C implies that the SW dynamics has discrete time entropy decay with rate $\delta = 1/C$, and therefore, by Lemma 2.2, satisfies $T_{\text{mix}} = O(\log n)$. Theorem 1.8 from the introduction now follows immediately.

Proof of Theorem 1.8. For the Potts model one has $e^{\beta\Delta} = O(1/b)$. Therefore, the results follows from Theorem 5.1, Lemma 1.17, Theorem 6.1 and [BCP$^+$21, Lemma 1.8]. □

Let $\{V_1, ..., V_k\}$ be the k-partition of G, where $k \le \Delta + 1$, as in Section 3. For all $j \in [k]$ let $\nu(\cdot \,|\, \sigma_{V_j^c}, A)$ denote the measure ν conditioned on $\sigma_{V_j^c} = \{\sigma_v,\, v \notin V_j\}$ and $A \subset E$. We use $\text{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)$ for the corresponding conditional entropy and $\nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)\right]$ for its expectation with respect to ν. Theorem 6.1 will follow from the following lemmas.

Lemma 6.2. *For all $f : \Omega_{\text{j}} \mapsto \mathbb{R}_+$ and all $j \in [k]$ we have*

$$\nu\left[\text{Ent}_\nu(f \,|\, A)\right] \ge \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)\right].$$

Lemma 6.3. *There exists a constant $\delta_1 > 0$ such that, for all $f : \Omega_{\text{j}} \mapsto \mathbb{R}_+$ and all $j \in [k]$,*

$$\nu\left[\text{Ent}_\nu(f \,|\, \sigma)\right] + \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)\right] \ge \delta_1\, \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V_j^c})\right].$$

The constant δ_1 satisfies $1/\delta_1 = O(\beta\Delta e^{\beta\Delta})$.

Lemma 6.4. *If μ satisfies the k-partite factorization with constant C_{par}, then for all $f : \Omega_{\text{j}} \mapsto \mathbb{R}_+$,*

$$\sum_{j=1}^{k} \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V_j^c})\right] \ge \delta_2 \text{Ent}_\nu(f),$$

where $\delta_2 = \frac{1}{C_{\text{par}}}$.

Proof of Theorem 6.1. By combining the bounds from Lemmas 6.2, 6.3 and 6.4 we get

$$\nu\left[\text{Ent}_\nu(f \,|\, \sigma) + \text{Ent}_\nu(f \,|\, A)\right] \ge \frac{\delta_1\delta_2}{k}\text{Ent}_\nu(f), \tag{6.3}$$

and so, using also $k \le \Delta + 1$, the spin/edge factorization holds with constant

$$C = \frac{k}{\delta_1\delta_2} = C_{\text{par}} \times O(\beta\Delta^2 e^{\beta\Delta}).$$ □

We turn to the proof of Lemmas 6.2, 6.3 and 6.4. In the special case of bipartite graphs these correspond to Lemmas 4.3, 4.4 and 4.5 in [BCP$^+$21], respectively. For Lemmas 6.2 and 6.4 the adaptation to our setting is straightforward. The proof of Lemma 6.3, the core of the argument, requires some modification. The main difference with the proof in [BCP$^+$21] is in the definition of the measures $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ below, since in the bipartite case one only needs to consider the measures $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ for $x \in V_j$ while here one needs to define $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ for all $x \in V$. Once this is taken care of, however, the proof proceeds essentially in the same way.

Proof of Lemma 6.2. This is an instance of the same monotonicity already seen in Lemma 2.4. In this particular case, it follows from the argument in the proof of Lemma 4.3 in [BCP$^+$21] by simply substituting σ_O with $\sigma_{V_j^c}$ in that proof. □

Proof of Lemma 6.3. Let us fix $j \in [k]$, and an arbitrary ordering of the independent sets $V_1, \dots, V_k$, such that V_j is the lowest independent set, that is $V_j < V_i$ for all $i \neq j$. We use xy to denote the edge $\{x, y\}$, and view the edge configuration A as a vector in $\{0,1\}^E$. Clearly, if $xy \in E$ then $x \in V_i$ and $y \in V_\ell$ for some $i \neq \ell$. For any $x \in V$ we write $N(x)$ for the set of neighbors of x which belong to a higher independent set, that is if $x \in V_i$ then $y \in N(x)$ iff $xy \in E$ and $y \in V_\ell$ for some $V_\ell > V_i$. Note that, since V_j is the lowest independent set, if $x \in V_j$ then $N(x)$ coincides with the set of all neighbors of x. The main observation here is that, by definition of the measure ν, for any fixed configuration $\sigma_{V_j^c}$ of spins on V_j^c, the conditional probability $\nu(\cdot \,|\, \sigma_{V_j^c})$ is a product measure

$$\nu(\cdot \,|\, \sigma_{V_j^c}) = \bigotimes_{x \in V} \nu_x(\cdot \,|\, \sigma_{V_j^c}), \tag{6.4}$$

where the single measures $\nu_x(\cdot \,|\, \sigma_{V_j^c})$, $x \in V$, are described as follows. For each $x \in V_j$, $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ is the law on $\{1, \dots, q\} \times \{0,1\}^{N(x)}$ obtained by picking the spin of site x according to the Potts measure on x conditioned on the spin of its neighbors in V_j^c and then, independently for every $y \in N(x)$ with $\sigma_x = \sigma_y$ by taking A_{xy} a Bernoulli(p) random variable, and for every $y \in N(x)$ with $\sigma_x \neq \sigma_y$ by setting $A_{xy} = 0$. For $x \in V_j^c$ instead, the single measure $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ is the law on $\{1, \dots, q\} \times \{0,1\}^{N(x)}$ obtained by taking a Dirac mass on $\{1, \dots, q\}$ according to the assigned spin value σ_x, and such that independently for every $y \in N(x)$ with $\sigma_x = \sigma_y$, A_{xy} is a Bernoulli(p) random variable, and for every $y \in N(x)$ with $\sigma_x \neq \sigma_y$ one has $A_{xy} = 0$. Note that, by construction, if $x \in V_j^c$ then the spins σ_x, σ_y, for $y \in N(x)$, are all assigned once we condition on $\sigma_{V_j^c}$.

The measure $\nu(\cdot \,|\, \sigma_{V_j^c}, A)$, obtained by further conditioning on a valid configuration of all edge variables A compatible with the fixed spins $\sigma_{V_j^c}$, is again a product measure:

$$\nu(\cdot \,|\, \sigma_{V_j^c}, A) = \bigotimes_{x \in V} \nu_x(\cdot \,|\, \sigma_{V_j^c}, A), \tag{6.5}$$

where $\nu_x(\cdot \,|\, \sigma_{V_j^c}, A)$ is defined as follows. If $x \in V_j$, $\nu_x(\cdot \,|\, \sigma_{V_j^c}, A)$ is the probability measure on $\{1, \dots, q\} \times \{0,1\}^{N(x)}$ that is uniform in the spin variable if x has no incident edges in A, and is concentrated on the unique admissible value given $\sigma_{V_j^c}$ and A otherwise, and it is a Dirac mass in the edge variables according to A. If $x \in V_j^c$, $\nu_x(\cdot \,|\, \sigma_{V_j^c}, A)$ is a Dirac mass on $\{1, \dots, q\} \times \{0,1\}^{N(x)}$ according to the assigned spin value σ_x and edge variables A.

Next, we note that $\nu(\cdot \,|\, \sigma)$ is a product of Bernoulli(p) random variables over all monochromatic edges in σ, while it is concentrated on $A_{xy} = 0$ on all remaining edges. Therefore we may write

$$\nu(\cdot \,|\, \sigma) = \bigotimes_{x \in V} \nu_x(\cdot \,|\, \sigma), \tag{6.6}$$

where $\nu_x(\cdot \,|\, \sigma)$ is the probability measure on $\{1, \dots, q\} \times \{0,1\}^{N(x)}$ given by a Dirac mass at the assigned spin σ_x and the product of Bernoulli(p) variables on all edges xy such that $y \in N(x)$ and $\sigma_x = \sigma_y$, and a Dirac mass at $A_{xy} = 0$ if $y \in N(x)$ and $\sigma_x \neq \sigma_y$.

We write $\mathrm{Ent}_x(\cdot \,|\, \sigma_{V_j^c})$, $\mathrm{Ent}_x(\cdot \,|\, \sigma_{V_j^c}, A)$, $\mathrm{Ent}_x(\cdot \,|\, \sigma)$ for the entropies with respect to the distributions $\nu_x(\cdot \,|\, \sigma_{V_j^c})$, $\nu_x(\cdot \,|\, \sigma_{V_j^c}, A)$, $\nu_x(\cdot \,|\, \sigma)$ respectively. The next key observation is that, for every site x, there is a local factorization of entropies in the following sense. There exists a constant $\delta_1 \in (0,1]$ such that $1/\delta_1 = O(\beta\Delta e^{\beta\Delta})$, and such that for all functions $f \geq 0$ and all σ and $x \in V$,

$$\nu_x\left[\mathrm{Ent}_x(f \,|\, \sigma) \,|\, \sigma_{V_j^c}\right] + \nu_x\left[\mathrm{Ent}_x(f \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c}\right] \geq \delta_1 \,\mathrm{Ent}_x(f \,|\, \sigma_{V_j^c}). \tag{6.7}$$

In the case $x \in V_j$ this follows exactly as in Lemma 4.7 from [BCP$^+$21] for bipartite graphs and is thus omitted. If instead $x \in V_j^c$ then, recalling that by construction $\nu_x(\cdot \,|\, \sigma_{V_j^c})$ is a Dirac mass on the spin value at x and a product measure on the edge variables at xy, $y \in N(x)$, one has $\nu_x(\cdot|\sigma_{V_j^c}) = \nu_x(\cdot|\sigma)$ and therefore

$$\mathrm{Ent}_x(f \,|\, \sigma_{V_j^c}) = \mathrm{Ent}_x(f \,|\, \sigma) = \nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma) \,|\, \sigma_{V_j^c} \right],$$

and thus δ_1 can be taken to be 1 in this case.

Next, we would like to lift inequality (6.7) to the product measure $\nu(\cdot \,|\, \sigma_{V_j^c}) = \otimes_{x \in V} \nu_x(\cdot \,|\, \sigma_{V_j^c})$. Let $x = 1, \dots, n$ denote an arbitrary ordering of the sites $x \in V$. For all $x \in V$ we let $A_x \in \{0,1\}^{N(x)}$ be the random variable corresponding to the state of the edges xy such that $y \in N(x)$. We write $\xi_x = (\sigma_x, A_x)$ for the pair of variables corresponding to any $x \in V$. Note that, under the conditional distribution $\nu(\cdot \,|\, \sigma_{V_j^c})$, the random variables ξ_x, $x \in V$, are independent. Thus, we may write

$$\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}) = \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}) \,|\, \sigma_{V_j^c} \right], \tag{6.8}$$

where $g_x = \nu \left[f \,|\, \sigma_{V_j^c}, \xi_{x+1}, \dots, \xi_n \right]$, $g_0 = f$ and $g_n = \nu \left[f \,|\, \sigma_{V_j^c} \right]$. This identity is an instance of the decomposition in Lemma 2.3.

Putting together (6.7) and (6.8) yields

$$\begin{aligned} \delta_1 \,\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}) &\le \sum_{x=1}^{n} \nu \left[\nu_x \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_{V_j^c} \right] + \nu_x \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right] \,|\, \sigma_{V_j^c} \right] \\ &= \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) + \mathrm{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right]. \end{aligned} \tag{6.9}$$

To conclude the proof we can now proceed exactly as in the proof of Lemma 4.8 from [BCP$^+$21]. We obtain the following two inequalities:

$$\begin{aligned} \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_{V_j^c} \right] &\le \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_{V_j^c} \right], \\ \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right] &\le \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right]. \end{aligned}$$

These two inequalities combined with (6.9) yield that

$$\delta_1 \,\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}) \le \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_{V_j^c} \right] + \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right]. \tag{6.10}$$

The desired result follows by taking expectations with respect to ν in (6.10). □

Proof of lemma 6.4. From the definition of conditional entropy as well as the fact that $\nu(\cdot \,|\, \sigma_{V_j}, \sigma_{V_j^c}) = \nu(\cdot \,|\, \sigma)$, we get

$$\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}) = \mathrm{Ent}_\nu \left(\nu \left[f \,|\, \sigma \right] \,|\, \sigma_{V_j^c} \right) + \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_{V_j^c} \right]. \tag{6.11}$$

(see eq. (4.5), (4.6) from Lemma 4.5 in [BCP$^+$21]). Now, since the function $\nu \left[f \,|\, \sigma \right]$ depends only on the spin configuration σ, one has the identity

$$\sum_{j=1}^{k} \nu \left[\mathrm{Ent}_\nu(\nu[f \,|\, \sigma] \,|\, \sigma_{V_j^c}) \right] = \sum_{j=1}^{k} \mu \left[\mathrm{Ent}(\nu[f \,|\, \sigma] \,|\, \sigma_{V_j^c}) \right], \tag{6.12}$$

where the entropy in the right hand side is with respect to μ and not with respect to ν. Since k-partite factorization holds by assumption,

$$\sum_{j=1}^{k} \mu\left[\mathrm{Ent}(\nu[f\,|\,\sigma]\,|\,\sigma_{V_j^c})\right] \geq \delta_2\,\mathrm{Ent}\left(\nu\left[f\,|\,\sigma\right]\right), \tag{6.13}$$

where $\delta_2 = 1/C_{\mathrm{par}}$. By taking functions depending only on σ_{V_j} for a single V_j one easily sees that C_{par} must be at least 1. Then, taking expectation and summing over j in (6.11), and combining with (6.12) and (6.13), we get

$$\sum_{j=1}^{k} \nu\left[\mathrm{Ent}_\nu(f\,|\,\sigma_{V_j^c})\right] \geq \delta_2\,\mathrm{Ent}_\nu\left(\nu\left[f\,|\,\sigma\right]\right) + k\,\nu\left[\mathrm{Ent}_\nu(f\,|\,\sigma)\right].$$

Using the simple decomposition $\mathrm{Ent}_\nu(f) = \mathrm{Ent}_\nu\left(\nu\left[f\,|\,\sigma\right]\right) + \nu\left[\mathrm{Ent}_\nu(f\,|\,\sigma)\right]$, and the fact that $\delta_2 \leq 1 \leq k$, we conclude that

$$\sum_{j=1}^{k} \nu\left[\mathrm{Ent}_\nu(f\,|\,\sigma_{V_j^c})\right] \geq \delta_2\,\mathrm{Ent}_\nu(f). \qquad \square$$

References

[Ale13] A. Alexanderian. On continuous dependence of roots of polynomials on coefficients. Technical notes, 2013.

[ALO20] 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. MR4232133

[BCLM11] F. Barthe, D. Cordero-Erausquin, M. Ledoux, and B. Maurey. Correlation and Brascamp-Lieb inequalities for Markov semigroups. *International Mathematics Research Notices*, 10:2177–2216, 2011. MR2806562

[BCP+21] 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. MR4398940

[BCV20] 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. MR4060350

[BD97] 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.

[BGP16] 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. MR3432570

[BN19] 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. MR4019887

[BT06] S. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. MR2283379

[Cap04] P. Caputo. Spectral gap inequalities in product spaces with conservation laws. *Stochastic analysis on large scale interacting systems*, 39:53–88, 2004. MR2073330

[CC09] E. A. Carlen and D. Cordero-Erasquin. Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities. *Geometric and Functional Analysis*, 19:373–405, 2009. MR2545242

[CCL03] E. A. Carlen, M. C. Carvalho, and M. Loss. Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta Mathematica*, 191:1–54, 2003. MR2020418

[CDM+19] 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. MR3909603

[CDPP09] P. Caputo, P. Dai Pra, and G. Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. In *Annales de l'IHP Probabilités et statistiques*, volume 45, pages 734–753, 2009. MR2548501

[Ces01] 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. MR1853483

[CGM21] 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. MR4203344

[CGvV21] 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. MR4262527

[CHV20] E. Csóka, V. Harangi, and B. Virág. Entropy and expansion. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 2428–2444. Institut Henri Poincaré, 2020. MR4164843

[CLL04] E. A. Carlen, E. H. Lieb, and M. Loss. A sharp analog of Young's inequality on S^N and related entropy inequalities. *The Journal of Geometric Analysis*, 14:487–520, 2004. MR2077162

[CLV20] 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. MR4232132

[CLV21] 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. MR4398939

[CM03] P. Caputo and F. Martinelli. Relaxation time of anisotropic simple exclusion processes and quantum Heisenberg models. *The Annals of Applied Probability*, 13(2):691–721, 2003. MR1970283

[CMT15] 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. MR3434252

[Con22] G. Conforti. A probabilistic approach to convex (φ)-entropy decay for Markov chains. *The Annals of Applied Probability*, 32(2):932–973, 2022. MR4414692

[CP21] P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388:793–818, 2021. MR4334247

[DGJ09] 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. MR2498672

[DPPP02] P. Dai Pra, A. M. Paganoni, and G. Posta. Entropy inequalities for unbounded spin systems. *The Annals of Probability*, 30(4):1959–1976, 2002. MR1944012

[DS96] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 9(3):695–750, 1996. MR1410112

[EHMT17] M. Erbar, C. Henderson, G. Menz, and P. Tetali. Ricci curvature bounds for weakly interacting Markov chains. *Electronic Journal of Probability*, 22:1–23, 2017. MR3646066

[ELL17] 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. MR3726607

[ES88] 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. MR0965465

[FGYZ21] 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. MR4262528

[FK72] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model I. Introduction and relation to other models. *Physica*, 57(4):536–564, 1972. MR0359655

[Gri06] G. R. Grimmett. *The Random-Cluster Model*, volume 333. Springer-Verlag, Berlin, 2006. MR2243761

[GSS19] F. Götze, H. Sambale, and A. Sinulis. Higher order concentration for functions of weakly dependent random variables. *Electronic Journal of Probability*, 24, 2019. MR4003138

[Häg96] O. Häggström. The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104(2):231–253, 1996. MR1373377

[Hay06] 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.

[HS07] 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. MR2326236

[HS19] J. Hermon and J. Salez. Modified log-Sobolev inequalities for strong-Rayleigh measures. *Preprint, arXiv:1902.02775*, 2019.

[Jer95] 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. MR1369061

[Led99] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités XXXIII*, pages 120–216, 1999. MR1767995

[Liu21] K. Liu. From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk. In *Proceedings of APPROX-RANDOM*, 2021. MR4366587

[LP17] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, 2017. MR3726904

[Mar99] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics*, pages 93–191. Springer, 1999. MR1746301

[Mar19] K. Marton. Logarithmic Sobolev inequalities in discrete product spaces. *Combinatorics, Probability & Computing*, 28(6):919–935, 2019. MR4015662

[Mey00] C. D. Meyer. *Matrix analysis and applied linear algebra*, volume 71. SIAM, 2000. MR1777382

[MO94] 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. MR1269387

[MS13] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013. MR3059200

[Oll09] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009. MR2484937

[RR19] 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. MR4019886

[Sal21] J. Salez. Cutoff for non-negatively curved Markov chains. *Preprint, arXiv:2102.05597*, 2021.

[SS20] H. Sambale and A. Sinulis. Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli*, 26(3):1863–1890, 2020. MR4091094

[SW87] R. H. Swendsen and J. S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.

[SZ92] 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. MR1152374

[Ull14] M. Ullrich. Rapid mixing of Swendsen-Wang dynamics in two dimensions. *Dissertationes Mathematicae*, 502:1–65, 2014. MR3222829

[Vig00] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000. MR1757969

[Vil21] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Society, 2021. MR1964483