ROMA TRE
UNIVERSITÀ DEGLI STUDI

CORSO DI DOTTORATO DI RICERCA IN MATEMATICA

XXXV CICLO DEL CORSO DI DOTTORATO

Entropy and mixing time of non-local Markov chains and non-linear recombination models

Candidato:
Daniel Parisi Firma:_____

Relatore:
Prof. Pietro Caputo Firma:_____

Coordinatore:
Prof. Alessandro Giuliani Firma:_____

Summary

This thesis is divided in two parts. In the first part we consider spin systems in the d-dimensional lattice $\mathbb{Z}^d$ satisfying the so-called strong spatial mixing condition (SSM), a standard condition corresponding to exponential decay of correlations with distance between spins on the lattice. We show that the relative entropy functional of the corresponding Gibbs measure satisfies a family of inequalities which control the entropy on a given region $V \subset \mathbb{Z}^d$ in terms of a weighted sum of the entropies on blocks $A \subset V$ when each A is given an arbitrary nonnegative weight α_A. These inequalities generalize the well known logarithmic Sobolev inequality for the Glauber dynamics. Moreover, they provide a natural extension of the classical Shearer inequality satisfied by the Shannon entropy. Finally, they imply a family of modified logarithmic Sobolev inequalities which give tight bounds on the mixing time of arbitrary weighted block dynamics of heat bath type. Then, we show that the tools involved in the analysis of the block dynamics can be adapted to the study of the mixing time of the Swendsen-Wang dynamics for the ferromagnetic Ising and Potts models on the integer lattice $\mathbb{Z}^d$. This dynamics is a widely used Markov chain that has largely resisted sharp analysis because it is non-local, i.e., it changes the entire configuration in one step. In particular we prove that, whenever SSM holds, the mixing time on any n-vertex cube in $\mathbb{Z}^d$ is $O(\log n)$, and we prove this is tight by establishing a matching lower bound on the mixing time. The previous best known bound was $O(n)$. The proof of this fact utilizes a new factorization of the entropy in the joint probability space over spins and edges that underlies the Swendsen-Wang dynamics, which extends to general bipartite graphs of bounded degree. Our result then follows from the fact that this factorization implies the modified log-Sobolev inequality. This factorization also leads to several additional results, including mixing time bounds for a number of natural local and non-local Markov chains on the joint space, as well as for the standard random-cluster dynamics.

We finally extend our analysis to spin systems on an arbitrary graph $G = (V, E)$ with finite spin space, in which case we prove that a contractive cou-

pling 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. Roughly speaking, a distribution is spectrally independent if the maximum eigenvalues of the influence matrices associated to the distribution and its conditional distributions are upper bounded. 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.

The second part is devoted to the study of a nonlinear recombination model from population genetics as a combinatorial version of the Kac-Boltzmann equation from kinetic theory. Following Kac's approach, the nonlinear model is approximated by a mean field linear evolution with a large number of particles. In our setting, the latter takes the form of a generalized random transposition dynamics. Our main results establish a uniform in time propagation of chaos with quantitative bounds, and a tight entropy production estimate for the generalized random transpositions, which holds uniformly in the number of particles. As a byproduct of our analysis we obtain sharp estimates on the speed of convergence to stationarity for the nonlinear equation, both in terms of relative entropy and total variation norm.

The first part of this thesis is based on [27, 12, 11], while the second part is based on [28].

Acknowledgments

I would like to express my deeply gratitude to my supervisor Prof. Pietro Caputo for his consistent support and guidance during my three years of PhD.

Furthermore I would like to thank all the research collaborators for the hard work and the insightful discussions: Antonio Blanca, Pietro Caputo, Zongchen Chen, Alistair Sinclair, Daniel Stefankovic, Eric Vigoda.

Finally, I would like to thank my friends and my family, for all the support they have given to me.

Contents

List of Figures

Part I

Entropy factorization and mixing time of non-local Markov chains

Chapter 1

Introduction

1.1 Spin systems and Gibbs samplers

1.1.1 Spin systems

Spin systems arise in many scientific fields such as statistical physics and theoretical computer science. Given a graph $G = (V, E)$ and a fixed integer $q \geq 2$, a spin system on the graph G is a Gibbs distribution μ on the space of configurations $\Omega = S^V$, which contains terms that depend on the spin values at each vertex and at each pair of adjacent vertices. S is the set of spins, and a spin configuration $\sigma \in \Omega$ is an assignment of spins $\sigma : V \to S$ to each vertex of G. When $A \subset V$, σ_A is the projection $\sigma_A : A \to S$ defined as $\sigma_A(x) = \sigma(x)\ \forall x \in A$. When S is finite it is often indicated as $S := [q]$, where we define $[q] := \{1, \dots, q\}, q \geq 2$. A *boundary condition* (or *pinning*) τ for the spin system is a fixed assignment of spins to the boundary of V. We write μ^τ to denote the Gibbs distribution μ with boundary condition τ. We refer to section 2.1 for formal definitions.

The Potts model is one of the most common and widely studied example of spin system. It is defined by the following distribution

$$\mu(\sigma) = \frac{\exp(\beta M(\sigma))}{Z(G, \beta)},$$

where $M(\sigma) = \sum_{xy \in E} \mathbf{1}(\sigma_x = \sigma_y)$ is the set of monochromatic edges, $Z(G, \beta)$ is the partition function and $\beta \in \mathbb{R}$ is the model parameter, which captures the strength of interactions between the spins; in mechanical statistics β is associated to the inverse temperature of the model. When $\beta \geq 0$ the model is said to be ferromagnetic, if $\beta < 0$ it is called antiferromagnetic. When $q = 2$ this spin system is known as the Ising model.

Other well known examples of spin systems are the q-colorings model, de-

fined by

$$\mu(\sigma) = \frac{\mathbf{1}(\sigma \in \Omega_{G,q})}{|\Omega_{G,q}|},$$

where $\Omega_{G,q} := \{\sigma \in \Omega : \sigma \text{ is a proper coloring of } G\}$, and, for $q = 2$, the hardcore model, which is defined by

$$\mu(\sigma) = \frac{\lambda^{|\sigma|}\mathbf{1}(\sigma \in \mathcal{I})}{Z(G,\lambda)},$$

where $\mathcal{I} = \{\text{independent sets of } G\}$ and the parameter $\lambda > 0$ is called the fugacity of the model.

One of the main tasks regarding the study of spin systems is to address the problem of sampling from the Gibbs distribution. As an example of this, let us consider the Hardcore model with fugacity 1 $\mu(\sigma) = \frac{1}{|\mathcal{I}|}\mathbf{1}(\sigma \in \mathcal{I})$, where $\mathcal{I}$ is the set of all independent sets of G; in this case directly sampling from this distribution would mean computing $|\mathcal{I}|$, which is a very hard problem to solve since $|\mathcal{I}|$ can be exponential in the number of vertices of the graph.

A classical and much simpler way to address the problem of sampling is to introduce sophisticated algorithms such that after a certain amount of time they approximately generate samples of the Gibbs distribution.

Monte Carlo Markov chains (MCMC) are standard and widely used Markov chains for sampling. Namely in every step of these algorithms, the current configuration will be randomly updated under appropriate rules, such that the distribution will eventually converge to its stationary measure, namely the measure preserved by the chain. The number of steps required to be sufficiently close to the stationary measure is called the mixing time of the chain, see Eq. (2.3.1), and the goal is to understand how the mixing time grows as the size of the graph increases.

1.1.2 Gibbs samplers: Block dynamics and Swendsen-Wang dynamics

In this thesis we are mostly interested in a particular class of MCMC, the Gibbs samplers. One of the main Gibbs sampler we consider is the block dynamics, that we define as follows. Let us fix a probability measure $\alpha = \{\alpha_A, A \subset V\}$ over 2^V (i.e. the set of all subsets of V) and consider the Markov chains that at each step picks a $A \subset V$ with a probability α_A and update its spins σ_A according to the

conditional distribution $\mu_A^{\sigma_{A^c}} = \mu(\cdot|\sigma_{A^c})$. The transition matrix is

$$P_\alpha(\sigma, \sigma') = \sum_{A \subset V} \alpha_A \mathbf{1}(\sigma_{A^c} = \sigma'_{A^c}) \mu_A^{\sigma_{A^c}}(\sigma'_A) \quad \forall \sigma, \sigma' \in \Omega, \tag{1.1.1}$$

and it is not hard to see that it is ergodic (i.e. aperiodic and irreducible), reversible with respect to μ (which in turn implies that μ is its stationary distribution) and that the Dirichlet form is defined as

$$D_\alpha(f, g) = \langle f, (1 - P_\alpha) g \rangle_\mu = \sum_{A \subset V} \alpha_A \mu \left[\mathrm{Cov}_A(f, g) \right] \quad \forall f, g \in L_2(\mu), \tag{1.1.2}$$

where $\mathrm{Cov}_A(f, g) := \mu_A[fg] - \mu_A[f]\mu_A[g]$ denotes the covariance of f, g with respect to μ_A. This Markov chain is called the $\alpha-$ weighted (heat bath) block dynamics, and one of our main contributions is to provide tight mixing time bounds for this chain when the spin system satisfies some decay of correlations properties.

When $G \subset \mathbb{Z}^d$ we will require that the spin systems satisfy a decay of correlation property that belongs to the family of *strong spatial mixing* conditions, and we will refer to it as SSM. In the case of finite spins it is one of many equivalent conditions introduced by Dobrushin and Shlosman [49] to characterize the so-called *complete analyticity* regime. This property, roughly speaking, expresses the fact that the correlation between spins at different vertices decreases exponentially with the distance between them. More precisely, given $V \subset \mathbb{Z}^d$ and a pair of fixed configurations ψ and ψ_x on the boundary of V such that ψ and ψ_x differ only in the spin of the vertex x, we say that the property $\mathcal{C}(V, K, a)$ is satisfied with constants $K, a > 0$ if the effect on the conditional distribution at a set $B \subset V$ decays as $K \exp(-ad(x, B))$. We then say that the spin system satisfies $SSM(K, a)$ if there exists $K, a > 0$ such that the property $\mathcal{C}(V, K, a)$ holds for any $V \subset \mathbb{Z}^d$. As we will see it is important to consider a relaxed spatial mixing condition that requires the decay to hold only for all sufficiently “fat" sets, in which case we denote the corresponding spatial mixing property as $SSM_L(K, a)$. It is important to remark that the latter definition is weaker than SSM, in the sense that there are spin systems that do not satisfy the condition $SSM(K, a)$ but for which the relaxed condition $SSM_L(K, a)$ holds if L is a suitably large constant, see remark 2.2.4. We refer to section 2.2.1 for a formal definition of SSM.

When G is an arbitrary graph with maximum degree Δ independent on $|V|$ and $S = [q]$, we will assume that the spin systems satisfy spectral independence, a powerful new approach for proving fast convergence of Markov chain Monte Carlo (MCMC) algorithms. A distribution μ is said to be $\eta-$spectrally independent if the maximum eigenvalue of the influence matrix J is less or equal than η

for all possible pinnings, where J is the $V \times [q] \times V \times [q]$ matrix defined as

$$J(x, a; y, b) := \mu(\sigma_y = b | \sigma_x = a) - \mu(\sigma_y = b), \quad x \neq y.$$

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 [4], for any 2-spin antiferromagnetic spin system in the tree uniqueness region [41], and for colorings [40, 132] it matched the best known parameter bounds using other algorithmic approaches. Moreover, recent work of Chen et al. [42] shows that spectral independence implies optimal mixing of the Glauber dynamics in all of these cases. We refer to section 2.2.2 for a formal definition of spectral independence.

One of the most known and extensively studied Gibbs sampler for the distribution μ is the Glauber dynamics. Note that this chain is a particular case of the more general block dynamics since it corresponds to the case where $\alpha_A = \frac{1}{|V|}\mathbf{1}(|A| = 1)$. Despite the vast literature concerning the study of mixing time of the Glauber dynamics, very few results have been proved for the more general $\alpha-$ weighted block dynamics. One of the main motivations for its study is to go beyond the locality of the Glauber dynamics, and thus being also able to provide optimal bounds for non-local Markov chains, i.e. chains that at each step updates the spins of large subsets of V, which may mix faster than local dynamics. An example of non local block dynamics is the following. If the graph has maximum degree Δ, then it is not hard to see that it is $k-$partite with $k \leq \Delta + 1$, so there exist k independent sets $V_1, \ldots, V_k \subset V$ such that $\bigsqcup_{i=1}^{k} V_i = V$. The $k-$partite block dynamics is then defined by the weights $\alpha_A = \frac{1}{k}\mathbf{1}(A = V_j, \exists\, j \in [k])$. If the graph is bipartite, then this chain is called the even-odd block dynamics.

Another popular Markov chain for sampling is the *Swendsen-Wang (SW)* dynamics [124], which utilizes the random-cluster representation of the Potts model to derive a sophisticated *non-local* Markov chain in which every vertex can update its spin in each step. From the current spin configuration $\sigma(t) \in \Omega$, the SW dynamics generates $\sigma(t+1) \in \Omega$ as follows:

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

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

3. For each connected component $\mathcal{C}$ in the subgraph $(V, A(t))$, independently choose a spin $s_{\mathcal{C}}$ uniformly at random from $[q]$ and assign $s_{\mathcal{C}}$ to all vertices

in $\mathcal{C}$, yielding $\sigma(t+1) \in \Omega$.

The Swendsen-Wang dynamics is ergodic, and has the Gibbs distribution as its stationary distribution; see [56] for a proof. This non-local dynamics has the ability to flip large regions of spins in one step and was thus originally proposed as an alternative algorithm for overcoming the slow convergence at low temperatures of the Glauber dynamics. We will show that the techniques and tools involved in the analysis of the mixing time of the block dynamics can be effectively extended and applied to the analysis of the mixing time of the SW dynamics.

1.2 Functional inequalities

Functional inequalities such as the Poincaré and the logarithmic Sobolev inequality have long played a key role in the analysis of convergence to equilibrium for spin systems. 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., [46, 95, 17].

For the Glauber dynamics associated to a lattice Gibbs measures in the high temperature regime, rather conclusive results were obtained around thirty years ago [78, 134, 123, 122, 90, 97]. Broadly speaking, the main results of these works can be summarized with the statement that for finite or compact spin space, if the spin system satisfies a spatial mixing condition, then the *relative entropy* functional of the Gibbs measure μ_V describing the system on any region $V \subset \mathbb{Z}^d$, satisfies an *approximate tensorization* of the form

$$\mathrm{Ent}_V f \leq C \sum_{x \in V} \mu_V[\mathrm{Ent}_x f]\,, \tag{1.2.1}$$

where $C \geq 1$ is a constant, f is a nonnegative function, and $\mathrm{Ent}_V f$ is the relative entropy

$$\mathrm{Ent}_V f = \mu_V\left[f \log\left(f/\mu_V f\right)\right],$$

with $\mathrm{Ent}_x f$ denoting the local entropy at x, a function of all spins except for the spin at vertex x. The key feature of this inequality is its dimensionless character, namely the fact that the constant $C \geq 1$ is independent of both the region V, and the boundary condition fixed in $\mathbb{Z}^d \setminus V$, which we have omitted from our notation for simplicity. The value $C = 1$ is attained in the trivial case of independent spins. The papers mentioned above formulate their results in terms of *logarithmic Sobolev inequalities*, but we find it natural to restate them in terms of the tensorization inequality (1.2.1), which seems to have a more fundamental character in our

setting. Anyhow, if the spin space is finite, the statement (1.2.1) is equivalent to the standard logarithmic Sobolev inequality for the single site heat bath Markov chain, see e.g. [26, 101].

The proof of these results was obtained through refined recursive techniques, which exploit the spatial mixing assumption to establish some form of factorization of the entropy functional. We refer to the surveys [95, 70] for systematic expositions of these techniques. A particularly simple and effective approach was later developed in [36] and [45], who independently showed that the spatial mixing condition implies a factorization estimate of the form

$$\mathrm{Ent}_V f \leq (1+\varepsilon)\,\mu_V[\mathrm{Ent}_A f + \mathrm{Ent}_B f]\ , \tag{1.2.2}$$

where A, B are e.g. two overlapping rectangular regions in $\mathbb{Z}^d$, with $V = A \cup B$, and $\varepsilon > 0$ is a constant that can be made suitably small provided the overlap between A and B is sufficiently thick. Here $\mathrm{Ent}_A f$ denotes the relative entropy of f with respect to the Gibbs measure μ_A and it is thus a function of all spins outside of the region A. If the inequality (1.2.2) is available, then a relatively simple recursion leads to the desired conclusion (1.2.1).

While the inequality (1.2.1) is well suited for the analysis of the single site heat bath Markov chain, it is not very helpful in the analysis of more general block dynamics. With that motivation in mind, we address the question of the validity of a version of the inequality (1.2.1) where single sites $x \in V$ are replaced by arbitrary blocks $A \subset V$. Let $\alpha = (\alpha_B)_{B \subset V}$ be an arbitrary probability distribution over subsets of V, and define the minimum "coverage probability" of a vertex by

$$\delta(\alpha) = \min_{x \in V} \sum_{B : B \ni x} \alpha_B. \tag{1.2.3}$$

General block factorization of entropy holds with constant C if for all weights α, for all $f : \Omega \to \mathbb{R}_+$:

$$\delta(\alpha)\,\mathrm{Ent}_V\, f \leq C \sum_{B \subset V} \alpha_B\, \mu[\mathrm{Ent}_B\, f]. \tag{1.2.4}$$

Approximate entropy tensorization (1.2.1) 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 (1.2.4) holds with $C = 1$, in which case it is known as the Shearer inequality; see lemma 2.4.4.

In this thesis we will show that (1.2.4) holds with a constant independent on $|V|$. However, what we generally aim for is the best constant such that general block factorization holds, which is independent of $|V|$ in the cases we consider,

but it may not be in other cases, see [20] for a recent interesting result for permutations where the constant C is explicitly calculated when $\alpha_A = \binom{|V|}{l}^{-1} \mathbf{1}(|A| = l)$ and it is optimally dependent on $|V|$.

Important progress was obtained recently in [13] concerning the linearized version of (1.2.4). Namely, if we replace the entropy functional $\mathrm{Ent}_V f$ by the *variance* functional

$$\mathrm{Var}_V f = \mu_V \left[(f - \mu_V f)^2 \right],$$

then (1.2.4) becomes the Poincaré inequality

$$\delta(\alpha) \, \mathrm{Var}_V f \leq C \sum_{A \subset V} \alpha_A \, \mu_V [\mathrm{Var}_A f], \tag{1.2.5}$$

which we may refer to as the *block factorization of variance*. Notice that the inequality (1.2.5) provides the lower bound $\delta(\alpha)/C$ on the spectral gap of the *α-weighted block dynamics*.

One of the main results of [13] shows that, if the system satisfies the strong spatial mixing assumption, then it must satisfy the special case of (1.2.5) where the weights α are all either zero or one, but otherwise arbitrary, and where $\delta(\alpha)$ is replaced by the indicator $\mathbf{1}_{\delta(\alpha)>0}$, see [13, Theorem 1.2]. The proofs in [13] however rely crucially on coupling arguments as in [55], which do not seem to apply directly to prove the stronger statement (1.2.4).

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 2.4.6) and constitutes a key concept in the proof of the theorems below.

1.3 Main results for spin systems in $\mathbb{Z}^d$

We establish the block factorization of entropy, namely the full statement (1.2.4), for nearest neighbour spin systems satisfying the strong spatial mixing assumption. For instance, it will follow that the block factorization of entropy holds throughout the whole one phase region for the ferromagnetic Ising/Potts models in two dimensions, provided V in (1.2.4) is a sufficiently regular set in the sense of [96], see Section 2.2.1.

As a corollary, we obtain estimates on the speed of convergence to equilibrium of any block dynamics. Indeed, Jensen's inequality shows that, for any $A \subset V \subset$

$\mathbb{Z}^d$,

$$\mathrm{Ent}_A f \leq \mathrm{Cov}_A(f, \log f),$$

and therefore (1.2.4) implies the following *modified logarithmic Sobolev inequality* for any α-weighted block dynamics:

$$\delta(\alpha)\,\mathrm{Ent}_V f \leq C\,\mathcal{D}_{V,\alpha}(f, \log f). \tag{1.3.1}$$

In particular, the block factorization of entropy implies the exponential decay in time of the relative entropy, with rate at least $\delta(\alpha)/C$, for any α-weighted block dynamics, see remark 2.3.4. Moreover, if the spin state is finite the bound (1.3.1) implies the upper bound

$$T_{\mathrm{mix}}(V, \alpha) \leq D\,\delta(\alpha)^{-1} \log |V|, \tag{1.3.2}$$

where $|V|$ is the cardinality of the set V, D is some new absolute constant and $T_{\mathrm{mix}}(V, \alpha)$ denotes the total variation *mixing time* of the α-weighted block dynamics, see Lemma 2.4.6. We remark that (1.3.2) provides tight bounds on the mixing time for a large class of non-local Markov chains, for which previously known estimates were only polynomial in the size of V. A particularly interesting example, that is often used in the Monte Carlo Markov Chain practice, is the case of the even/odd chain. In this case (1.3.2) gives a tight $O(\log |V|)$ bound whereas the best previously known estimate was of order $|V|$, see [13].

If the spin state is finite it is also possible to use (1.2.4) to derive a standard logarithmic Sobolev inequality for the α-weighted block dynamics in the form

$$\mathrm{Ent}_V f \leq s(\alpha)\ \mathcal{D}_{V,\alpha}\Big(\sqrt{f}, \sqrt{f}\Big)\,, \tag{1.3.3}$$

with the constant

$$s(\alpha) = D\,\delta(\alpha)^{-1} \max_{A:\,\alpha_A > 0} \log(1/\mu_{A,*}),$$

where D is an absolute constant and $\mu_{A,*}$ is the minimum value attained by the probability measure μ_A, minimized over the choice of the implicit boundary condition in $\mathbb{Z}^d \setminus A$. We note that (1.3.3) contains as a special case the well known logarithmic Sobolev inequality for the single site heat bath Markov chain.

Theorem 1.3.1. *Suppose that the spin system satisfies $SSM(K, a)$ for some constants $K, a > 0$, and let $\mathbb{F}$ be the set of all subsets of $\mathbb{Z}^d$. Then there exists a constant $C > 0$ such that for all $V \in \mathbb{F}$, $\tau \in \Omega_{V^c}$, for all nonnegative weights $\alpha = \{\alpha_A,\ A \subset V\}$, for all*

$f : \Omega_V \mapsto \mathbb{R}_+$ with $f \log^+ f \in L^1(\mu_V^\tau)$,

$$\delta(\alpha)\,\mathrm{Ent}_V^\tau f \le C \sum_{A\subset V} \alpha_A\, \mu_V^\tau[\mathrm{Ent}_A f]\,, \tag{1.3.4}$$

where $\delta(\alpha) = \min_{x\in V} \sum_{A:\, A\ni x} \alpha_A$. If instead the spin system satisfies $SSM_L(K,a)$ for some constants $K, a > 0$, $L \in \mathbb{N}$, then the conclusion (1.3.4) *continues to hold, provided we require that $V \in \mathbb{F}^{(L)}$, where $\mathbb{F}^{(L)}$ is the set of all subsets of $\mathbb{Z}^d$ that are finite unions of lattice cubes of side L.*

Note that we are not assuming the spin space to be finite here. Theorem 1.3.1 has the following corollary for the α-weighted block dynamics defined by (1.1.1). Below, $\mathcal{D}^\tau_{V,\alpha}(f,g)$ denotes the Dirichlet form (1.1.2) evaluated at a given boundary condition $\tau \in \Omega_{V^c}$.

Corollary 1.3.2. *If the spin system satisfies $SSM(K,a)$ for some constants $K, a > 0$, then the following modified logarithmic Sobolev inequalities hold: for all $V \in \mathbb{F}$, all $\tau \in \Omega_{V^c}$, for all weights α, for all $f : \Omega_V \mapsto \mathbb{R}_+$ with $f \log^+ f \in L^1(\mu_V^\tau)$,*

$$\delta(\alpha)\,\mathrm{Ent}_V^\tau f \le C\, \mathcal{D}^\tau_{V,\alpha}(f, \log f), \tag{1.3.5}$$

where $\delta(\alpha)$ and C are the same constants appearing in (1.3.4)*. In particular, if the spin state S is finite, then there exists a constant $D > 0$ such that for all $V \in \mathbb{F}$, $\tau \in \Omega_{V^c}$, for all weights α, the mixing time $T^\tau_{\mathrm{mix}}(V,\alpha)$ of the Markov chain with Dirichlet form $\mathcal{D}^\tau_{V,\alpha}$ satisfies*

$$T^\tau_{\mathrm{mix}}(V,\alpha) \le D\, \delta(\alpha)^{-1} \log |V|. \tag{1.3.6}$$

Moreover, if the spin state is finite, then $SSM(K,a)$ implies the following logarithmic Sobolev inequalities: there exists a constant $D > 0$ such that for all $V \in \mathbb{F}$, all $\tau \in \Omega_{V^c}$, for all weights α, all $f \ge 0$,

$$\begin{aligned} \mathrm{Ent}_V^\tau f &\le s(\alpha)\ \mathcal{D}^\tau_{V,\alpha}\Big(\sqrt{f}, \sqrt{f}\Big)\,, \\ s(\alpha) &= D\, \delta(\alpha)^{-1} \max_{A:\, \alpha_A>0} \log(1/\mu_{A,*}), \end{aligned} \tag{1.3.7}$$

where

$$\mu_{A,*} = \min_{\tau\in\Omega_{A^c}} \min_{\sigma_A\in\Omega_A:\, \mu_A^\tau(\sigma_A)>0} \mu_A^\tau(\sigma_A)\,.$$

Finally, all statements above continue to hold if we only assume $SSM_L(K,a)$ for some constants $K, a > 0$ and $L \in \mathbb{N}$, provided we restrict to $V \in \mathbb{F}^{(L)}$.

Corollary 1.3.2 is a straightforward consequence of Theorem 1.3.1. Indeed, the modified log-Sobolev inequality (1.3.5) follows from the block factorization

(1.3.4) via Jensen's inequality. Moreover, the bound (1.3.6) is a standard consequence of (1.3.5), see e.g. [46, 17] and Lemma 2.4.6. Finally, (1.3.7) follows immediately from (1.3.4) and a well known bound comparing $\mathrm{Ent}_A f$ to $\mathrm{Var}_A\sqrt{f}$, see [46, Corollary A.4].

The analysis of the block dynamics provide tools and techniques useful to provide interesting results for the SW dynamics. The most important result shows that the mixing time of the SW dynamics is $O(\log |V|)$ whenever SSM holds, and this is tight. The notion of SSM used for the results concerning the SW dynamics is restricted to $n-$ vertex cubes, even though it is possible to generalize it to more general regions of $\mathbb{Z}^d$, for example our results for the SW dynamics holds if we assume SSM_L, as pointed out in remark 4.1.5.

Theorem 1.3.3. *In an n-vertex cube of $\mathbb{Z}^d$, for all integer $q \geq 2$, SSM implies that for all boundary conditions $T_{\text{mix}}(SW) = \Theta(\log n)$.*

In the presence of a boundary condition, we consider the Gibbs distribution on V conditional on the assignment τ on the boundary of V. The case where there is no boundary condition is known as the *free boundary* case and is also covered by our results.

Therefore, we obtain the following immediate corollary of Theorem 1.3.3.

Corollary 1.3.4. *In an n-vertex square region of $\mathbb{Z}^2$, for all $q \geq 2$, all $\beta < \beta_c(q)$ and all boundary conditions, we have $T_{\text{mix}}(SW) = \Theta(\log n)$.*

The best previous bound in the setting of Corollary 1.3.4 was $T_{\text{mix}}(SW) = O(n)$ and follows from the results in [13]. Nam and Sly [110] recently proved an $O(\log n)$ mixing time bound (as well as the cutoff phenomenon) for the periodic boundary condition for sufficiently high temperatures ($\beta \ll \beta_c(q)$), a stronger assumption than SSM. In higher dimensions $d \geq 3$, SSM is not known to hold up to the corresponding uniqueness threshold (it is only known for sufficiently small β; see [95]), but we expect the SW dynamics to be rapidly mixing throughout the high temperature regime for all $d \geq 3$.

The key to our improved mixing time analysis is a novel *factorization of entropy* based on the joint probability space of spins and edges that underlies the SW dynamics, see section 4.1.1 for a precise definition. This factorization implies that the relative entropy decays at a constant rate, which in turn implies a tight bound on the mixing time via a modified log-Sobolev inequality. In contrast, previous bounds for the SW dynamics [127, 16, 71, 13, 14] have used the spectral gap, which inherently loses a factor of $O(n)$ when transferred to mixing time bounds and cannot deliver a tight result.

A priori the correct order of the mixing time of the SW dynamics is unclear. In some settings, such as on the complete graph (the mean-field Potts model) for

all β below the uniqueness threshold, the dynamics mixes in $\Theta(1)$ steps [89, 63]. To complement our main result of an $O(\log n)$ upper bound, we also establish a lower bound of $\Omega(\log n)$ for all boundary conditions whenever SSM holds.

Our methods also provide new results for the low-temperature regime $\beta > \beta_c(q)$ in $\mathbb{Z}^2$ for specific boundary conditions. We say that a boundary condition τ is *monochromatic* if τ fixes the spin of every boundary vertex to the same color. One of the most fundamental open problems in the study of the Glauber dynamics for the Ising and Potts model concerns the mixing time at low temperatures with a monochromatic boundary [99, 91]. We provide new bounds for the mixing time of the SW dynamics in this setting.

Theorem 1.3.5. *In an n-vertex square region of $\mathbb{Z}^2$, for all $q \geq 2$ and all $\beta > \beta_c(q)$ we have $T_{\text{mix}}(SW) = O(n \log n)$ for the free or monochromatic boundary condition.*

The best previously known bound for the mixing time of the SW dynamics in an n-vertex square region of $\mathbb{Z}^2$ when $\beta > \beta_c(q)$ was $O(n^2 \log^2 n)$, which follows from the results in [16, 127]; see also [94] for better (sub-linear) bounds for the mixing time when $q = 2$ and $\beta \gg \beta_c(q)$. The bound in Theorem 1.3.5 is likely not tight, and establishing that the SW dynamics mixes in $O(\log n)$ steps in $\mathbb{Z}^2$ throughout the low-temperature regime remains an important open problem. Furthermore, our result for low temperature with a monochromatic boundary does not extend to higher dimensions $d \geq 3$, since it crucially uses the self-duality for the associated random-cluster distribution on $\mathbb{Z}^2$. For the Ising model with $d \geq 3$, the state of the art seems to be the results from [94] for $\beta \gg \beta_c(q)$.

1.4 Main results for spin system on a general graph

We now present our main results for general spin systems. On one side, 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 tools from the previous results. On the other side, we show 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.

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 [21]. 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.

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. 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 [58] and Remark 1.4.7). We refer to [23, 60, 44, 119] for further relevant works on the relations between curvature and entropy in Markov chains.

1.4.1 Consequences of spectral independence

Now we state the main implications of spectral independence for spin systems defined on an arbitrary $n-$vertex graph G with maximum degree Δ. In this setting, we also require that the spin system is $b-$marginally bounded, that is the marginal probability for any vertex is lower bounded by a constant b, see definition 2.2.7.

Theorem 1.4.1. *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* (1.2.4) *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 2.4.6, one recovers Theorem 1.9 in [42] as a special case of the above result. As another example, for a bipartite graph, Theorem 1.4.1 implies $O(\log n)$ mixing time of the even-odd dynamics.

When α is the uniform distribution over all subsets of a given size ℓ, we refer to (1.2.4) as the *ℓ-uniform block factorization of entropy* or ℓ-UBF for short. In [42], an important step in the proof of Theorem 1.9 is establishing ℓ-UBF with $\ell \sim \theta n$ for some $\theta \in (0, 1)$. To prove Theorem 1.4.1 for arbitrary blocks we establish that ℓ-UBF implies general block factorization of entropy; see Theorem 5.1.2 for a detailed statement.

Following the approach presented in the previous section and using our general result in Theorem 1.4.1, 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 1.4.9 below.

Theorem 1.4.2. *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 $C = \left(\frac{2}{b}\right)^{O\left(\Delta(1+\frac{\eta}{b})\right)}$.*

1.4.2 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, see section 5.3.1 for a formal definition.

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

Theorem 1.4.3.

(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 $\eta \le \frac{2\gamma^2}{\epsilon}$.

Note a κ-contractive coupling for the Hamming distance immediately implies $O(n \log n)$ mixing time of the Glauber dynamics (see, e.g., [21, 86]). 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 [21], yields a mixing time bound which depends on this ratio. Second, as discussed in the previous theorems, spectral independence (together with the marginal boundedness) implies optimal bounds on the mixing time and entropy decay rate for arbitrary heat-bath block dynamics.

We can extend Theorem 1.4.3 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 5.3.3.

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 (5.3.8) for the precise definition of D.

Theorem 1.4.4. *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 $\eta = \frac{2\gamma^2 DM}{1-\kappa}$.*

Remark 1.4.5. An important and rather straightforward consequence of Theorem 1.4.4 is that it shows a connection between SSM and spectral independence. Namely, if $G \subset \mathbb{Z}^d$ and the spin system is finite, then SSM implies spectral independence with $\eta = O(1)$, so that all the results regarding the block dynamics and the SW dynamics in $\mathbb{Z}^d$ stated in section 1.3 are a particular case of Theorems 1.4.1 and 1.4.9. This follows because SSM implies the existence of a contractive coupling for the select-update dynamics known as the tiled bath block dynamics, see [13] and lemma 3.1 therein.

Theorem 1.4.4 generalizes Theorem 1.4.3(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 5.3.11. See also Theorem 5.3.9 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 1.4.4 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 1.4.3 we obtain that the Dobrushin uniqueness condition implies spectral independence. The Dobrushin uniqueness condition is a classical condition in statistical physics which holds if the maximum column of the sum of the Dobrushin dependency matrix

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

is at most $1-\epsilon$ for some $\epsilon > 0$, where $\mathcal{S}_{x,y}$ is the set of all pairs of configurations on $V \setminus \{y\}$ that can differ only at x.

$R(x,y)$ 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 1.4.3 we prove that the Dobrushin uniqueness condition implies spectral independence. Moreover, our result holds under generalizations of the Dobrushin uniqueness condition. Hayes [76] generalized it to the following spectral condition: if $\|R\|_2 \leq 1 - \epsilon$ for some $\epsilon > 0$, then the mixing time of the Glauber dynamics is $O(n \log n)$. This was further generalized by Dyer et al. [54] 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 5.3.5 for a more detailed discussion.

Theorem 1.4.6. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \leq 1 - \epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

Previously, Marton [102] (see also [67, 121]) showed that the spectral condition in Theorem 1.4.6 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 1.4.7. 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 [111]. Combining Theorem 1.4.3 with Theorem 1.4.1 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 (1.2.3). Replacing Theorem 1.4.3 with its generalization Theorem 1.4.4 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. [58].

1.4.3 Applications of the main results

We finally discuss a few examples of applications of our results. 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 [79] 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 [129] and further improved to $q > (\frac{11}{6} - \epsilon_0)\Delta$ for some small $\epsilon_0 \approx 10^{-5} > 0$ by Chen et al. [38] 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 [77], and also obtain optimal bounds on the log-Sobolev and modified log-Sobolev constants.

Theorem 1.4.8. *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 [108] 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 [124]. In [14], 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 before 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 [76, 128, 19] 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, [19] 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 5.3.17 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 [74] and it behaves as $\beta_u = \frac{\ln q}{\Delta - 1} + O(1)$; see [19]. 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 1.4.9. *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 $\beta < \max\{\beta_0, \beta_1\}$.*

1.5 Organization

The organization of the first part of the thesis goes as follows. In chapter 2 we give the formal definitions of spin systems and the corresponding decay of correlation properties. We also gather a few well known facts that we use throughout the thesis. In the next two chapters we prove the results stated in section 1.3. In particular, in chapter 3 we prove Theorem 1.3.1, while in chapter 4 we prove all the results concerning SW dynamics, namely Theorems 1.3.3 and 1.3.5 stated in section 1.3 of the introduction, and we also discuss further results implied by the spin/edge factorization concerning other dynamics. Chapter 5 is devoted to the analysis of spin systems under spectral independence and establishing spectral independence. We sketch the proofs of Theorems 1.4.1 and 1.4.2 stated in section 1.4 of the introduction and their implications. We then turn our analysis to establishing spectral independence, and we prove all the results in section 1.4.2 of the introduction. In section 5.4 we also discuss a reformulation of some of the key facts that were derived in [42] in the setting of spin systems through a recursive scheme.

Chapter 2

Preliminaries

2.1 Spin systems and Gibbs distributions

2.1.1 Spin systems with finite spin space

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 $A_{xy} \in \mathbb{R}_+^{q \times q}$, $\{x, y\} \in E \cup \partial E$, 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 $\sigma \in \Omega$,

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

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. We 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 q-spin system.

The Ising/Potts model. 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 Hardcore 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. The q-colorings model corresponds to $A_{xy}(a, a') = \mathbf{1}(a \neq a')$ and $B_x(a) = 1$. This distribution corresponds to the limit $\beta \to -\infty$ in the Potts model.

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.

2.1.2 Spin systems in $\mathbb{Z}^d$

In our analysis in $\mathbb{Z}^d$ it is convenient to use the notations in [27] which include the case where the spin space is not finite; this is important because we are able to prove that the block factorization statement in $\mathbb{Z}^d$ also holds when the spin space S is not finite. The underlying graph is the d-dimensional integer lattice $\mathbb{Z}^d$, with vertices $x = (x_1, \ldots, x_d)$, and edges E defined as unordered pairs xy of vertices x and y such that $\sum_{i=1}^d |x_i - y_i| = 1$. We call $d(\cdot, \cdot)$ the resulting graph distance. For any set of vertices $\Lambda \subset \mathbb{Z}^d$, the exterior boundary is $\partial\Lambda = \{y \in \Lambda^c : d(y, \Lambda) = 1\}$, where $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$. We write $\mathbb{F}$ for the set of finite subsets $\Lambda \subset \mathbb{Z}^d$.

We take the single spin state to be an arbitrary probability space $(S, \mathscr{S}, \nu)$. Given any region $\Lambda \subset \mathbb{Z}^d$, the associated configuration space is the product space $(\Omega_\Lambda, \mathcal{F}_\Lambda) = (S^\Lambda, \mathscr{S}^\Lambda)$, whose elements are denoted by $\sigma_\Lambda = \{\sigma_x,\, x \in \Lambda\}$ with $\sigma_x \in S$ for all x. The *apriori measure* on Ω_Λ is the product measure $\nu_\Lambda = \otimes_{x\in\Lambda}\nu_x$.

Given a *bounded* measurable symmetric function $U : S \times S \mapsto \mathbb{R}$, the pair potential, and a *bounded* measurable function $W : S \mapsto \mathbb{R}$, the single site potential, for any $\Lambda \in \mathbb{F}$, and $\tau \in \Omega_{\Lambda^c}$, the Hamiltonian $H^\tau_\Lambda : \Omega_\Lambda \mapsto \mathbb{R}$ is defined by

$$H^\tau_\Lambda(\sigma_\Lambda) = -\sum_{\substack{xy\in E:\\ x,y\in\Lambda}} U(\sigma_x,\sigma_y) - \sum_{\substack{xy\in E:\\ x\in\Lambda, y\in\partial\Lambda}} U(\sigma_x,\tau_y) - \sum_{x\in\Lambda} W(\sigma_x).$$

The Gibbs measure in the region $\Lambda \in \mathbb{F}$ with boundary condition $\tau \in \Omega_{\Lambda^c}$ is the probability measure μ^τ_Λ on $(\Omega_\Lambda, \mathcal{F}_\Lambda)$ defined by

$$\mu^\tau_\Lambda(d\sigma_\Lambda) = \frac{1}{Z^\tau_\Lambda} \exp\left[-H^\tau_\Lambda(\sigma_\Lambda)\right] \nu_\Lambda(d\sigma_\Lambda)\,,$$

where Z^τ_Λ is the normalizing constant. The spin system is called *permissive* if for every $\Lambda \in \mathbb{F}$, for every $\tau \in \Omega_{\Lambda^c}$, there exists $\sigma_\Lambda \in \Omega_\Lambda$ with positive mass under μ^τ_Λ, that is such that $\mu^\tau_\Lambda(\sigma_\Lambda) > 0$. Well known examples of finite permissive spin systems include the *hard-core model* with parameter λ, for any $\lambda > 0$, and the uniform distribution over *proper q-colorings*, for any integer $q \geq 2d+1$.

A classical continuous spin system is obtained when S is a compact subset of $\mathbb{R}^n$ and ν is the uniform distribution over S. The O(n) *model*, for $n \geq 2$, corresponds to the case where S is the unit sphere in $\mathbb{R}^n$, $\beta \in \mathbb{R}$,

$$U(s,s') = \beta\langle s, s'\rangle\,, \qquad W(s) = \beta\langle s, v\rangle\,,$$

for some fixed vector $v \in S$, with $\langle\cdot,\cdot\rangle$ denoting the standard inner product in $\mathbb{R}^n$.

A permissive spin system is called *irreducible* if the single site heat bath Markov chain on Λ with boundary condition τ is irreducible for any choice of $\Lambda \in \mathbb{F}$ and $\tau \in \Omega_{\Lambda^c}$, see [13, Section 2]. Our main results in $\mathbb{Z}^d$ will apply to permissive irreducible spin systems.

Remark 2.1.1. The setup introduced above includes unbounded (continuous or discrete) spins. When $S = \mathbb{Z}_+$ for instance it covers the particle systems considered in [45]. It should be however clear that the boundedness assumptions on the interaction U rules out many interesting models in the unbounded setting.

Remark 2.1.2. Concerning possible extensions of Theorem 1.3.1 and Corollary 1.3.2 to more general settings, we remark that the definitions given above can be extended to include spatially non-homogeneous models, with pair potentials

U and site potentials W replaced by edge dependent functions U_{xy} and site dependent functions W_x respectively. It is not difficult to check that all results can be extended to include these cases provided that all the estimates involved in our assumptions are uniform with respect to the new potentials. Finally, we remark that our setup is restricted to the case of nearest neighbor interactions, and the extension of our main results to more general finite range spin systems is not immediate. Indeed, our proof makes explicit use of the nearest neighbor structure at various places. We believe however that a similar approach can be used, provided the decomposition into even and odd sites used in our proof is replaced by more general tilings such as the ones used in [13].

2.1.3 DLR property for spin systems

A fundamental feature of the family of measures $\{\mu_\Lambda^\tau,\, \Lambda \subset V, \tau \in \Omega_{\Lambda^c}\}$ is the so-called DLR property:

$$\mu_V \mu_\Lambda f = \mu_V f\,, \tag{2.1.1}$$

valid for all $\Lambda \subset V$, and for all bounded measurable function $f : \Omega_V \mapsto \mathbb{R}$.

2.2 Decay of correlations

2.2.1 Strong spatial mixing in $\mathbb{Z}^d$

The precise formulation we give here coincides with the one adopted in Cesi's paper [36]. For any $\Delta \subset \Lambda \in \mathbb{F}$ we call $\mu_{\Lambda,\Delta}^\tau$ the marginal of μ_Λ^τ on Ω_Δ. A version of the Radon-Nikodym density of $\mu_{\Lambda,\Delta}^\tau$ with respect to ν_Δ is given by the function

$$\psi_{\Lambda,\Delta}^\tau(\sigma_\Delta) := \frac{1}{Z_\Lambda^\tau} \int \exp\left[-H_\Lambda^\tau(\eta_{\Lambda\setminus\Delta}\sigma_\Delta)\right] \nu_{\Lambda\setminus\Delta}(d\eta_{\Lambda\setminus\Delta}),$$

where $\eta_{\Lambda\setminus\Delta}\sigma_\Delta$ denotes the configuration $\xi \in \Omega_\Lambda$ such that $\xi_x = \eta_x$ if $x \in \Lambda \setminus \Delta$ and $\xi_x = \sigma_x$ if $x \in \Delta$.

Definition 2.2.1. Given constants $K, a > 0$, and $\Lambda \in \mathbb{F}$ we say that condition $\mathcal{C}(\Lambda, K, a)$ holds if for any $\Delta \subset \Lambda$, for all $x \in \partial\Lambda$:

$$\sup_{\tau,\tau'} \left\| \frac{\psi_{\Lambda,\Delta}^{\tau'}}{\psi_{\Lambda,\Delta}^{\tau}} - 1 \right\|_\infty \le K\, e^{-a\, d(x,\Delta)}, \tag{2.2.1}$$

where $\tau, \tau' \in \Omega_{\Lambda^c}$ are such that $\tau_y = \tau'_y$ for all $y \neq x$, and $\|\cdot\|_\infty$ denotes the L^∞ norm. We say that the spin system satisfies $SSM(K, a)$ if $\mathcal{C}(\Lambda, K, a)$ holds for all

$\Lambda \in \mathbb{F}$.

As emphasized in [96] it is often important to consider a relaxed spatial mixing condition that requires $\mathcal{C}(\Lambda, K, a)$ to hold only for all sufficiently "fat" sets Λ. The latter is defined as follows.

Definition 2.2.2. Given $L \in \mathbb{N}$, let $Q_L = [0, L-1]^d \cap \mathbb{Z}^d$ be the lattice cube of side L located at the origin. For any $y \in \mathbb{Z}^d$, define the translated cube $Q_L(y) = Ly + Q_L$. Let $\mathbb{F}^{(L)}$ be the set of all $\Lambda \in \mathbb{F}$ of the form

$$\Lambda = \bigcup_{y \in \Lambda'} Q_L(y)$$

for some $\Lambda' \in \mathbb{F}$. The spin system satisfies $SSM_L(K, a)$ if $\mathcal{C}(\Lambda, K, a)$ holds for all $\Lambda \in \mathbb{F}^{(L)}$.

For systems without hard constraints it is well known that $SSM(K, a)$, for some K, a, is always satisfied in dimension one, and that for any dimension $d > 1$ it holds under the assumption of suitably high temperature, see e.g. [95]. It is important to note that the validity of both $SSM(K, a)$ and $SSM_L(K, a)$ can be ensured by checking finite size conditions only [93].

For the results concerning the SW dynamics in section 4, when μ is the Potts model, it is convenient to consider a spatial mixing condition that requires the condition $\mathcal{C}(\Lambda, K, a)$ to hold only for lattice cubes, see remark 4.1.5.

Definition 2.2.3. We say that *SSM with respect to cubes* holds if there exist $K, a > 0$ such that $\mathcal{C}(\Lambda, K, a)$ holds for every cube $\Lambda \subset \mathbb{Z}^d$.

In $\mathbb{Z}^2$, SSM is known to hold for all $q \geq 2$ and all $\beta < \beta_c(q)$, where $\beta_c(q) = \ln(1 + \sqrt{q})$ is the uniqueness threshold [9, 1, 98]. Namely, if $\beta < \beta_c(q)$ then there isn't a dominant spin class and there are only local correlations, whereas if $\beta > \beta_c(q)$ then there is a dominant spin class that stands out with respect to the others.

Remark 2.2.4. $SSM(K, a)$ can be strictly stronger than $SSM_L(K, a)$, that is there are spin systems that do not satisfy the condition $SSM(K, a)$ but for which the relaxed condition $SSM_L(K, a)$ holds if L is a suitably large constant. We refer to [96] for a thorough discussion of this subtle point. As a consequence of results in [98, 1, 9] it is also known that the two-dimensional ferromagnetic Potts model satisfies $SSM_L(K, a)$, for some $K, a > 0$ and $L \in \mathbb{N}$, throughout the whole one phase region, that is for all values of temperature and external field within the uniqueness region, there exist constants $K, a > 0$ and L such that $SSM_L(K, a)$ holds.

Finally, we note that $\mathcal{C}(\Lambda, K, a)$ is too strong a requirement in the case of systems with hard constraints, since $\mu^{\tau'}_{\Lambda,\Delta}$ may be not absolutely continuous with respect to $\mu^{\tau}_{\Lambda,\Delta}$. However, since (2.2.1) will only be relevant if $d(x, \Delta)$ is sufficiently large, in order to have a meaningful assumption for permissive spin systems with hard constraints, we may rephrase the condition $SSM_L(K, a)$ by requiring, for all $\Lambda \in \mathbb{F}^{(L)}$, that (2.2.1) holds for all $\Delta \subset \Lambda$ and $x \in \partial\Lambda$ such that $d(x, \Delta) \geq L/2$.

2.2.2 Spectral independence for $q-$spin systems

Let us consider a spin system with finite state space $[q]$. 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. We recall the definition of the set $\mathcal{X} = \{(x, a) : x \in V, a \in \Omega_x\}$ where $\Omega_x = \{\tau \in [q] : \exists \sigma \in \Omega, \sigma_x = \tau\}$.

Definition 2.2.5 (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 [4] 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 [40, 132] in different ways. Here we use the definition from [40] which is more suitable for our applications in Section 5.3 for establishing spectral independence, but we could also work with the definition from [132] 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 2.2.6 (Spectral independence)**.** We say that a spin system is *η-spectrally independent* if for all pinnings $\tau \in \mathcal{T}$ we have $\lambda_1(J^\tau) \leq \eta$.

We point out that SSM implies spectral independence in the case of $q-$spin systems, see remark 1.4.5.

2.2.3 Marginal boundedness

There is one additional property of the Gibbs distribution that will be relevant; 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 2.2.7 (Marginal boundedness)**.** A spin system is *b-marginally bounded* if for all pinnings τ, all $x \in V$, all $a \in \Omega_x^\tau$ we have $\mu^\tau(\sigma_x = a) \geq b$.

2.3 Mixing time and entropy

Here we state some well known general properties of the mixing time and the entropy. All the proofs of these properties are in the appendix. Let P be the transition matrix of an ergodic Markov chain with finite state space Ω and stationary distribution μ. Let $P^t(X_0, \cdot)$ denote the distribution of the chain after t steps starting from the initial state $X_0 \in \Omega$. The *mixing time* $T_{\text{mix}}(P)$ of the chain is defined as

$$T_{\text{mix}}(P) = \max_{X_0 \in \Omega} \min \left\{ t \geq 0 : \left\| P^t(X_0, \cdot) - \mu \right\|_{\text{TV}} \leq 1/4 \right\}, \tag{2.3.1}$$

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

We rely on functional inequalities related to entropy to bound the mixing time. The functionals below can be also defined for an arbitrary space Ω, but here we consider the finite case for simplicity. For a function $f : \Omega \mapsto \mathbb{R}$, let $\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$ and $\text{Var}_\mu(f) = \mu[f^2] - \mu[f]^2$ denote its mean and variance with respect to μ. Likewise, for $f : \Omega \to \mathbb{R}_+$, the *entropy* of f with respect to μ is defined as

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

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

A basic property of entropy that we shall use is the variational principle

$$\text{Ent}_\pi(f) = \sup \left\{ \pi[f\varphi], \ \pi[e^\varphi] \leq 1 \right\}, \tag{2.3.2}$$

valid for any distribution π, and any $f \geq 0$, where the supremum ranges over all functions $\varphi : \Omega \mapsto \mathbb{R}$ such that $\pi[e^\varphi] \leq 1$, see e.g. Proposition 2.2 in [85].

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

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

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

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

Definition 2.3.1. The pair (P,μ) satisfies the (standard) *log-Sobolev inequality* (LSI) with constant s if for all $f \geq 0$:

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

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

$$\mathcal{D}_P(f,\log f) \geq \varrho\,\mathrm{Ent}_\mu(f). \tag{2.3.3}$$

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

$$H(\nu P\,|\,\mu) \leq (1-\delta)H(\nu\,|\,\mu). \tag{2.3.4}$$

Remark 2.3.2. If ζ has density f with respect to μ (i.e., $\zeta = f\mu$), then ζP has density P^*f with respect to μ, where P^* is the *adjoint* or *time-reversal* matrix $P^*(\sigma,\sigma') = \frac{\mu(\sigma')}{\mu(\sigma)}P(\sigma',\sigma)$. Thus, (2.3.4) is equivalent to

$$\mathrm{Ent}_\mu(P^*f) \leq (1-\delta)\mathrm{Ent}_\mu(f), \tag{2.3.5}$$

for all $f \geq 0$ such that $\mu[f] = 1$. By homogeneity, this is equivalent to (2.3.5) for all $f \geq 0$. When P is reversible, that is when $P = P^*$, (2.3.4) is equivalent to $\mathrm{Ent}_\mu(Pf) \leq (1-\delta)\mathrm{Ent}_\mu(f)$ for all $f \geq 0$.

Inequality (2.3.4) can 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 2.3.3. *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 Ω is finite then (P,μ) satisfies the discrete time relative entropy decay with rate $\delta > 0$, then*

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

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

It is well known that the standard LSI with constant s implies entropy decay

in continuous time with rate $\delta = 2s$, since $\mathcal{D}_P(f, \log f) \geq 2\mathcal{D}_P(\sqrt{f}, \sqrt{f})$ for all $f \geq 0$, and this can be improved to $\delta = 4s$ in the reversible case; see [47, Lemma 2.7].

Remark 2.3.4. The Log-Sobolev inequality is equivalent to the so-called hypercontractivity (see [47, Theorem 3.5]), while the modified Log-Sobolev inequality (2.3.3) is equivalent to exponential decay of the relative entropy with rate δ for the continuous time kernel $K_t = e^{(P-1)t}$ (see [47, Theorem 3.6]). Note that we are not assuming reversibility. To see the relation between the MLSI and the entropy decay in continuous time, note that if $K_t = e^{(P-1)t}$ and f has mean $\mu[f] = 1$ then using $K_t^* = e^{(P^*-1)t}$ one checks that the time derivative of the relative entropy satisfies

$$\frac{d}{dt} H(\zeta K_t \,|\, \mu) = \frac{d}{dt} \mathrm{Ent}(K_t^* f) = -\mathcal{D}_P(K_t^* f, \log K_t^* f),$$

where $\zeta = f \cdot \mu$. Therefore (2.3.3) implies, for all $t \geq 0$:

$$H(\zeta K_t \,|\, \mu) \leq H(\zeta \,|\, \mu) e^{-\delta t}.$$

Here we recall a result of Miclo [104] showing in what sense the LSI implies the discrete time entropy decay.

Lemma 2.3.5. *If P is positive semi-definite and the pair (P^*P, μ) satisfies the standard LSI with constant s, then the discrete time entropy decay holds for (P, μ) with constant $\delta = s$. In particular, if P is reversible and (P, μ) satisfies the LSI with constant s, then for all $f \geq 0$:*

$$\mathrm{Ent}_\mu Pf \leq (1 - s)\mathrm{Ent}_\mu f.$$

2.4 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

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

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

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

The following lemma summarizes a standard decomposition of the entropy functional.

Lemma 2.4.1. *For any $\Lambda\subset V$, for any $f:\Omega\to\mathbb{R}_+$:*

$$\mathrm{Ent}(f) = \mu\left[\mathrm{Ent}_\Lambda(f)\right] + \mathrm{Ent}\left(\mu_\Lambda[f]\right). \tag{2.4.1}$$

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

$$\sum_{i=1}^{w}\mu\left[\mathrm{Ent}_{\Lambda_i}(\mu_{\Lambda_{i-1}}[f])\right] = \mu\left[\mathrm{Ent}_{\Lambda_w}(\mu_{\Lambda_0}[f])\right]. \tag{2.4.2}$$

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

Lemma 2.4.2. *For all $A\subset B\subset V$,*

$$\mu[\mathrm{Ent}_A(f)] \le \mu[\mathrm{Ent}_B(f)]. \tag{2.4.3}$$

Now we recall the definition of general block factorization of entropy.

Definition 2.4.3. 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,

$$\delta(\alpha)\,\mathrm{Ent}\,f \le C\sum_{A\subset V}\alpha_A\,\mu[\mathrm{Ent}_A\,f], \tag{2.4.4}$$

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

When μ_Λ is a product measure $\mu_\Lambda=\otimes_{x\in\Lambda}\mu_x$, then (2.4.4) is known as the Shearer inequality. We have the following lemma.

Lemma 2.4.4. *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 $f:\Omega\to\mathbb{R}_+$:*

$$\delta(\alpha)\,\mathrm{Ent}_\Lambda(f) \le \sum_{A\subset\Lambda}\alpha_A\,\mu_\Lambda[\mathrm{Ent}_A(f)], \tag{2.4.5}$$

that is μ_Λ satisfies the general block factorization of entropy with constant $C=1$. The same inequality holds for the variance functional, that is if we replace Ent *with* Var .

The following properties will also be used.

Lemma 2.4.5. *Let $\Lambda = A \cup B$ and assume that μ_Λ is a product $\mu_\Lambda = \mu_A \otimes \mu_B$. Then, for all $f \geq 0$:*

$$\mathrm{Ent}_\Lambda(\mu_B(f)) = \mu_\Lambda[\mathrm{Ent}_A(\mu_B(f))], \tag{2.4.6}$$

and for all $U \subset B$,

$$\mu_\Lambda[\mathrm{Ent}_A(\mu_B(f))] \leq \mu_\Lambda[\mathrm{Ent}_A(\mu_U(f))]. \tag{2.4.7}$$

2.4.1 Implications of block factorization

The following lemma relates the block factorization with the rate of the entropy decay of the block dynamics.

Lemma 2.4.6. *If the spin system is finite and 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. *$T_{\mathrm{mix}}(P_\alpha) \leq 1 + \frac{C}{\delta(\alpha)}[\log(8) + \log\log(1/\mu_*)]$, where $\mu_* = \min_{\sigma\in\Omega} \mu(\sigma)$.*

Chapter 3

Block factorization in $\mathbb{Z}^d$

In this chapter we prove the block factorization statement of Theorem 1.3.1. We start with a brief discussion of the main ideas involved in the proof. The proof starts with an observation already put forward in [13] for the case of the spectral gap, which allows us to reduce the general factorization problem to the problem of factorization with two special blocks only: the even sites and the odd sites. The latter is then analyzed via a recursion similar to that employed in Cesi's proof of (1.2.1), see [36]. The main obstacle in implementing the recursion here is the lack of an additive structure, which generates potentially large error terms when trying to restore a block from smaller components. To overcome this difficulty we develop a two-stage recursion, which combines a version of the two-block factorization estimate (1.2.2) together with a decomposition of the entropy which allows us to smear out the errors coming from the restoration of large blocks, see Theorem 3.3.6.

3.1 Some key tools

In this section we collect some key general facts that do not depend on the spatial mixing assumption. We start by proving a new general tensorization lemma. Then, we revisit the two-block factorization (1.2.2).

Some remarks on the notation are in order. We fix a region $V \in \mathbb{F}$ and a boundary condition $\tau \in \Omega_{V^c}$. To avoid heavy notation, we often omit explicit reference to V, τ. In particular, whenever possible we shall use the following shorthand notation

$$\mu f = \mu_V^\tau f\,, \qquad \mathrm{Ent} f = \mathrm{Ent}_V^\tau f\,. \tag{3.1.1}$$

Moreover, whenever we write μ_Λ or Ent_Λ for some $\Lambda \subset V$, we assume that the

implicit boundary condition outside Λ has been fixed, and it agrees with τ outside of V. Unless otherwise stated, f will always denote a nonnegative measurable function such that $f \log^+ f \in L^1(\mu)$. To avoid repetitions, we simply write $f \geq 0$ throughout. As a convention, we set $\mu_\emptyset f = f$ and $\mathrm{Ent}_\emptyset f = 0$.

3.1.1 A new tensorization lemma

Consider subsets

$$A_{i,j} \subset V \in \mathbb{F}\,, \qquad i = 1, \dots, n\,,\ j = 1, \dots, m,$$

such that $\cup_{i,j} A_{i,j} = \Lambda \subset V$, and define "row" subsets and "column" subsets:

$$R_i := \cup_{j=1}^m A_{i,j}\,, \qquad C_j := \cup_{i=1}^n A_{i,j}\,.$$

Assume that μ_Λ is a product measure along the partition $\{R_i,\, i = 1 \dots, n\}$ of Λ:

$$\mu_\Lambda = \otimes_{i=1}^n \mu_{R_i}\,.$$

Notice that this is the case if $\{R_1, \dots, R_n\}$ are such that $d(R_i, R_j) > 1$ for all $i \neq j$.

Lemma 3.1.1. *Let $s_i > 0$ be constants such that for each $i = 1, \dots, n$, for all $f \geq 0$,*

$$\mathrm{Ent}_{R_i} f \leq s_i \sum_{j=1}^m \mu_{R_i}[\mathrm{Ent}_{A_{i,j}} f]. \tag{3.1.2}$$

Then

$$\mathrm{Ent}_\Lambda f \leq s \sum_{j=1}^m \mu_\Lambda[\mathrm{Ent}_{C_j} f],$$

where $s = \max_i s_i$.

Proof. To simplify the notation, we write $\mu = \mu_\Lambda$ and $\mathrm{Ent}_\Lambda f = \mathrm{Ent} f$. Setting $\Lambda_k = \cup_{i=1}^k R_i$, with $\Lambda_0 = \emptyset$, from Lemma 2.4.1 we have

$$\mathrm{Ent} f = \sum_{k=1}^n \mu \left[\mathrm{Ent}_{\Lambda_k} \mu_{\Lambda_{k-1}} f\right].$$

Since μ_{Λ_k} is a product of μ_{R_i}, $i = 1, \dots, k$, we have

$$\mathrm{Ent} f = \sum_{k=1}^n \mu \left[\mathrm{Ent}_{R_k} \mu_{\Lambda_{k-1}} f\right].$$

From the assumption (3.1.2) we estimate

$$\mathrm{Ent} f \leq s \sum_{k=1}^{n} \sum_{j=1}^{m} \mu[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1}} f].$$

The proof is complete once we show that for each j,

$$\sum_{k=1}^{n} \mu[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1}} f] \leq \mu[\mathrm{Ent}_{C_j} f]. \tag{3.1.3}$$

Define $\Lambda_{k,j} = \Lambda_k \cap C_j$. From Lemma 2.4.1 we have

$$\mathrm{Ent}_{C_j} f = \sum_{k=1}^{n} \mu_{C_j} \left[\mathrm{Ent}_{\Lambda_{k,j}} \mu_{\Lambda_{k-1,j}} f \right].$$

For each j, k fixed, $\mu_{\Lambda_{k,j}}$ is a product of $\mu_{A_{i,j}}$, $i = 1, \ldots, k$. Hence,

$$\mathrm{Ent}_{C_j} f = \sum_{k=1}^{n} \mu_{C_j} \left[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f \right].$$

Therefore, (3.1.3) follows if we show that all j, k fixed:

$$\mu[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1}} f] \leq \mu \left[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f \right]. \tag{3.1.4}$$

To prove (3.1.4), notice that

$$\mu_{A_{k,j}} \mu_{\Lambda_{k-1}} f = \mu_{A_{k,j}} \mu_{\Lambda_{k-1}} \mu_{\Lambda_{k-1,j}} f = \mu_{\Lambda_{k-1}} \mu_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f,$$

where the second identity follows from the product structure $\mu_{\Lambda_k} = \otimes_{i=1}^{k} \mu_{R_i}$. Therefore,

$$\begin{aligned}
\mu \left[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1}} f \right] &= \mu \left[\mu_{\Lambda_{k-1}} f \log \left(\mu_{\Lambda_{k-1}} f / \mu_{A_{k,j}} \mu_{\Lambda_{k-1}} f \right) \right] \\
&= \mu \left[\mu_{\Lambda_{k-1}} \mu_{\Lambda_{k-1,j}} f \log \left(\mu_{\Lambda_{k-1}} \mu_{\Lambda_{k-1,j}} f / \mu_{\Lambda_{k-1}} \mu_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f \right) \right] \\
&= \mu \left[\mu_{\Lambda_{k-1,j}} f \log \left(\mu_{\Lambda_{k-1}} \mu_{\Lambda_{k-1,j}} f / \mu_{A_{k,j}} \mu_{\Lambda_{k-1}} \mu_{\Lambda_{k-1,j}} f \right) \right] \\
&\leq \mu \left[\mu_{A_{k,j}} \left(\mu_{\Lambda_{k-1,j}} f \log \left(\mu_{\Lambda_{k-1,j}} f / \mu_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f \right) \right) \right] \\
&= \mu \left[\mathrm{Ent}_{A_{k,j}} \mu_{\Lambda_{k-1,j}} f \right],
\end{aligned}$$

where the inequality follows from the variational principle (2.3.2). □

Here is an example to keep in mind, with n arbitrary and $m = 2$. Let us consider the following collection of subsets $\{R_1, \ldots, R_n\}$ where $R_i \in \mathbb{F}$ and $d(R_i, R_j) > 1$ for all $i \neq j$. Let $A_{i,1} = ER_i$ be the even sites in R_i and $A_{i,2} = OR_i$ be the odd

sites in R_i, where a vertex $x \in \mathbb{Z}^d$ is even or odd according to the parity of $\sum_{i=1}^d x_i$. Lemma 3.1.1 says that if we can factorize the even and odd sites on each R_i with some constant s_i, then we can also factorize, with the constant $\max_i s_i$, the even and odd sites on all $\Lambda = \cup_i R_i$. In this example, one has $A_{i,j} \cap A_{i,k} = \emptyset$ if $k \neq j$, so in particular $C_j \cap C_k = \emptyset$ for $k \neq j$, but it is interesting to note that this need not be the case in Lemma 3.1.1, that is each "row" R_i is allowed to be decomposed into arbitrary, possibly overlapping subsets $A_{i,j}$, $j = 1, \ldots, m$. We refer to Remark 3.2.3 for useful applications of the latter situation.

3.2 Preliminaries: two block factorizations

We shall need the following versions of an inequality of Cesi [36] and Dai Pra, Paganoni, Posta [45].

Lemma 3.2.1. *Take $A, B \in \mathbb{F}$ and $V = A \cup B$. Suppose that for some $\varepsilon \in (0,1)$:*

$$\|\mu_B \mu_A g - \mu g\|_\infty \leq \varepsilon\, \mu(|g|) \tag{3.2.1}$$

for all functions $g \in L^1(\mu)$. Then, for all functions $f \geq 0$,

$$\mathrm{Ent} f \leq \mu[\mathrm{Ent}_A f + \mathrm{Ent}_B f] + \theta(\varepsilon)\, \mathrm{Ent} f, \tag{3.2.2}$$

$$\mathrm{Ent} f \leq \mu[\mathrm{Ent}_A f + \mathrm{Ent}_B \mu_A f] + \theta(\varepsilon)\, \mathrm{Ent}\, \mu_A f, \tag{3.2.3}$$

where $\theta(\varepsilon) = 84\varepsilon(1-\varepsilon)^{-2}$.

Proof. The inequality (3.2.2) coincides with [36, Eq. (2.10)]. To prove (3.2.3) we use essentially the same argument. As in the proof of (3.2.2) we may restrict to the case where f is bounded, and bounded away from zero. Then

$$\begin{aligned}
\mathrm{Ent} f &= \mu\left[f \log\left(f/\mu_A f\right)\right] + \mu\left[f \log\left(\mu_A f/\mu f\right)\right] \\
&= \mu[\mathrm{Ent}_A f] + \mu\left[\mu_A f \log\left(\mu_A f/\mu f\right)\right] \\
&= \mu[\mathrm{Ent}_A f] + \mu[\mathrm{Ent}_B \mu_A f] + \mu\left[\mu_A f \log\left(\mu_B \mu_A f/\mu f\right)\right].
\end{aligned}$$

Cesi's inequality [36, Eq. (3.2)] says that the assumption (3.2.1) implies

$$\mu\left[f \log\left(\mu_B \mu_A f/\mu f\right)\right] \leq \theta(\varepsilon)\, \mathrm{Ent} f\,, \tag{3.2.4}$$

for all $f \geq 0$, where $\theta(\varepsilon) = 84\varepsilon(1-\varepsilon)^{-2}$. Therefore, the claim (3.2.3) follows from (3.2.4) applied with $\mu_A f$ in place of f. □

Remark 3.2.2. If μ is a product measure over A, B, that is $\mu = \mu_B \mu_A$, then one can take $\varepsilon = 0$ in Lemma 3.2.1. In this case (3.2.3) is actually an identity. In this sense (3.2.3) might be considered to be tighter than (3.2.2), although it is not true that $\mu[\mathrm{Ent}_B \mu_A f] \leq \mu[\mathrm{Ent}_B f]$ in the general non-product case: think for instance of some f which depends only on $A \setminus B$; in this case $\mu[\mathrm{Ent}_B f] = 0$ while it is possible that $\mu[\mathrm{Ent}_B \mu_A f] > 0$. For our purposes below it will be crucial to use both (3.2.2) and (3.2.3).

Remark 3.2.3. To appreciate the strength of the tensorization Lemma 3.1.1, consider a case where $V = \cup_{i=1}^n R_i$ with $R_i = A_i \cup B_i$ and suppose that μ_V is a product measure over the R_i's. If the condition (3.2.1) holds for every pair A_i, B_i, $i = 1, \ldots, n$, with the same constant $\varepsilon \in (0, 1)$, the combination of Lemma 3.2.1 and Lemma 3.1.1 shows that (3.2.2) holds uniformly in n, with $A = \cup_{i=1}^n A_i$ and $B = \cup_{i=1}^n B_i$. On the other hand, Lemma 3.2.1 alone seems unable to yield such a uniform estimate. Indeed, the assumption (3.2.1) does not tensorize: it is not hard to construct examples where (3.2.1) holds for every pair A_i, B_i, $i = 1, \ldots, n$, with the same error $\varepsilon \in (0, 1)$, but one has to take the error proportional to n in order to have (3.2.1) for $A = \cup_{i=1}^n A_i$ and $B = \cup_{i=1}^n B_i$. Here is a toy example of this phenomenon. Suppose $V = A \cup B$ with $A = \{1, \ldots, n\}$ and $B = \{n+1, \ldots, 2n\}$ and for all $i \in \{1, \ldots, n\}$, let $A_i = \{i\}$, $B_i = \{n+i\}$, and write σ_i and $\eta_i = \sigma_{n+i}$ for the spin at i and at $n+i$ respectively. Suppose that each spin takes values in $\{-, +\}$ and that the probability measure μ on $\{-, +\}^{2n}$ is defined by $\mu(\sigma, \eta) = \prod_{i=1}^n \mu_i(\sigma_i, \eta_i)$, where $\mu_i(\sigma_i, \eta_i) = \frac{1}{4}(1 + \varepsilon \sigma_i \eta_i)$, for some fixed $\varepsilon \in (0, 1)$. For any function $g : \{-, +\}^2 \mapsto \mathbb{R}$, for all $i \in \{1, \ldots, n\}$ one has

$$\mu_{B_i} \mu_{A_i} g - \mu_i g = \frac{\varepsilon \sigma_i}{4} \left[(1+\varepsilon)(g(+,+) - g(-,-)) + (1-\varepsilon)(g(-,+) - g(+,-)) \right].$$

Therefore $\|\mu_{B_i} \mu_{A_i} g - \mu_i g\|_\infty \leq \varepsilon \mu_i(|g|)$. On the other hand, the marginal on η of μ is the uniform distribution over $\{-, +\}^n$, so that if we choose $f(\sigma, \eta) = f(\eta) = \mathbf{1}_{\eta \equiv +}$ one has $\mu(|f|) = \mu(f) = 2^{-n}$, $\mu_A f = f$, and $\mu_B \mu_A f = 2^{-n} \prod_{i=1}^n (1 + \varepsilon \sigma_i)$. In particular, taking $\sigma \equiv +$ one finds

$$\|\mu_B \mu_A f - \mu f\|_\infty \geq 2^{-n} \left[(1+\varepsilon)^n - 1 \right] \geq \varepsilon n \mu(|f|).$$

3.3 Proof of the main results

We first reduce the general block factorization problem to the factorization into even and odd sites only.

3.3.1 Reduction to even and odd blocks

We partition the vertices of $\mathbb{Z}^d$ into even sites and odd sites, where x is *even* if $\sum_{i=1}^d x_i$ is an even integer, while x is *odd* if $\sum_{i=1}^d x_i$ is an odd integer. Given a set of vertices $V \in \mathbb{F}$ we write EV for the set of even vertices $x \in V$ and OV for the set of odd vertices $x \in V$. Whenever possible we simply write E for EV and O for OV. Notice that both μ_E and μ_O are *product measures*.

The reduction to even and odd blocks can be stated as follows. As usual we assume that a region $V \in \mathbb{F}$, and a boundary condition $\tau \in \Omega_{V^c}$ have been fixed, and we use the shorthand notation (3.1.1).

Proposition 3.3.1. *Suppose that for some constant $C > 0$ and some function $f \geq 0$,*

$$\mathrm{Ent} f \leq C\, \mu \left[\mathrm{Ent}_E f + \mathrm{Ent}_O f \right].$$

Then, for the same C and f, for all nonnegative weights $\alpha = \{\alpha_A,\, A \subset V\}$,

$$\delta(\alpha)\, \mathrm{Ent} f \leq 2\, C \sum_{A \subset V} \alpha_A\, \mu \left[\mathrm{Ent}_A f \right], \tag{3.3.1}$$

where $\delta(\alpha) = \min_{x \in V} \sum_{A : A \ni x} \alpha_A$.

Proposition 3.3.1 is a direct consequence of the Shearer's inequality, see Lemma 2.4.4.

Proof of Proposition 3.3.1. For any $A \subset V$, call EA and OA the even and odd sites in A respectively. Fix a choice of weights $\alpha = \{\alpha_A, A \subset V\}$. Since μ_E is a product measure on Ω_E, we may apply Lemma 2.4.4 with $\Lambda = E$ and weights α replaced by $\hat{\alpha} = \{\hat{\alpha}_U,\, U \subset E\}$, with $\hat{\alpha}_U = \sum_{A \subset V} \alpha_A \mathbf{1}_{EA = U}$. It follows that

$$\sum_{A \subset V} \alpha_A\, \mu_E[\mathrm{Ent}_{EA} f] \geq \delta_E(\alpha)\, \mathrm{Ent}_E f, \tag{3.3.2}$$

where $\delta_E(\alpha) = \min_{x \in E} \sum_{A : A \ni x} \alpha_A$. Similarly,

$$\sum_{A \subset V} \alpha_A\, \mu_O[\mathrm{Ent}_{OA} f] \geq \delta_O(\alpha)\, \mathrm{Ent}_O f, \tag{3.3.3}$$

with $\delta_O(\alpha) = \min_{x \in O} \sum_{A : A \ni x} \alpha_A$. Since $\delta_E(\alpha)$ and $\delta_O(\alpha)$ are both at least as large as $\delta(\alpha)$, the inequality (3.3.1) follows by summing (3.3.2) and (3.3.3), taking the expectation with respect to μ and noting that both $\mu[\mathrm{Ent}_{EA} f]$ and $\mu[\mathrm{Ent}_{OA} f]$ are at most $\mu[\mathrm{Ent}_A f]$. □

The rest of this section is concerned with the proof of the factorization into even and odd blocks. Namely, we prove the following theorem, which together

with Proposition 3.3.1 establishes the main result Theorem 1.3.1.

Theorem 3.3.2. *Suppose that the spin system satisfies $SSM(K,a)$ for some constants $K,a>0$. Then there exists a constant $C\geq 1$ such that for all $V\in\mathbb{F}$, $\tau\in\Omega_{V^c}$, for all $f\geq 0$,*

$$\mathrm{Ent}_V^\tau f\leq C\mu_V^\tau\left[\mathrm{Ent}_E f+\mathrm{Ent}_O f\right]. \tag{3.3.4}$$

If instead the spin system satisfies $SSM_L(K,a)$ for some constants $K,a>0$, $L\in\mathbb{N}$, then the same conclusion (3.3.4) *holds, provided we require that $V\in\mathbb{F}^{(L)}$.*

Remark 3.3.3. The constant C in (3.3.4) must be larger than 1, since if e.g. $f=f(\sigma_E)$ is a function depending only on the spins at even sites then the right hand side in (3.3.4) is equal to $\mu_V^\tau\left[\mathrm{Ent}_E f\right]\leq\mathrm{Ent}_V^\tau f$. From our proof one can in principle extract an explicit dependency of C on the various parameters defining the spin system and on the constants K,a in the strong mixing assumption. While this dependency is not optimal, one can for instance check, using a high temperature expansion, that in the limit of infinite temperature one recovers the constant $C=1$ corresponding to non interacting spins.

3.3.2 Proof of Theorem 3.3.2

The overall idea is to follow a recursive strategy based on a geometric construction introduced in [10], see also [36]. However, contrary to the problems studied in [10, 36], the error terms produced at each step of the iteration are too large in our setting to obtain directly the desired conclusion, see Theorem 3.3.6, and we will need an additional recursive argument to finish the proof, see Theorem 3.3.7. We first carry out the proof under the spatial mixing assumption $SSM(K,a)$, and then, in the end, consider the relaxed assumption $SSM_L(K,a)$.

Definition 3.3.4. Set $\ell_k=(3/2)^{k/d}$ and let $\mathbb{F}_k$ denote the set of all subsets $V\in\mathbb{F}$ such that, up to translation and permutation of the coordinates, V is contained in the rectangle

$$[0,\ell_{k+1}]\times\cdots\times[0,\ell_{k+d}].$$

Let $\delta(k)$ denote the largest constant $\delta>0$ such that

$$\delta\,\mathrm{Ent}_V^\tau f\leq\mu_V^\tau\left[\mathrm{Ent}_E f+\mathrm{Ent}_O f\right]$$

holds for all $V\in\mathbb{F}_k$, $\tau\in\Omega_{V^c}$, and all $f:\Omega_V\mapsto\mathbb{R}_+$.

Note that $\delta(k)\leq 1$ for any $k\in\mathbb{N}$, see Remark 3.3.3. On the other hand, the next lemma guarantees that it is positive for all $k\in\mathbb{N}$.

Lemma 3.3.5. *For every $k\in\mathbb{N}$, $\delta(k)>0$.*

Proof. If the spin system has no hard constraints one can use a perturbation argument from [78], see e.g. [26, Lemma 2.2] for the application to our setting. In particular, one obtains that there exists a constant $C > 0$ such that for all $k \in \mathbb{N}$:

$$\delta(k) \geq \exp\left(-C\ell_k^d\right).$$

In the presence of hard constraints, in the case of irreducible permissive systems one can argue as follows. It is known that any probability measure μ satisfies

$$\operatorname{Ent} f \leq C_0 \log(1/\mu_*) \operatorname{Var}\left(\sqrt{f}\right), \tag{3.3.5}$$

with $\mu_* = \min_\sigma \mu(\sigma)$, where the minimum is restricted to σ such that $\mu(\sigma) > 0$, and C_0 is an absolute constant, see [47, Corollary A.4]. Here Var denotes the variance functional of μ. For a finite permissive system in a region V one has $\mu_* \geq e^{-C|V|}$ for some $C > 0$ independent of V. Moreover, using the irreducibility assumption, a crude coupling argument shows that the spectral gap of the even/odd Markov chain is bounded away from zero in any fixed region $V \in \mathbb{F}$, see [13, Lemma 5.1]. In other words, for some constant $C_1 = C_1(k)$ one has

$$\operatorname{Var}(g) \leq C_1 \mu\left[\operatorname{Var}_E(g) + \operatorname{Var}_O(g)\right], \tag{3.3.6}$$

for any function g. Taking $g = \sqrt{f}$, the desired conclusion now follows from (3.3.5) and (3.3.6) using, for both μ_E and μ_O, the well known inequality $\operatorname{Var}(\sqrt{f}) \leq \operatorname{Ent} f$, which holds for any probability measure, see e.g. [84, Lemma 1].

□

Lemma 3.3.5 will be used as the base case for our induction.

Theorem 3.3.6. *Assume $SSM(K, a)$. There exists a constant $k_0 \in \mathbb{N}$ depending on K, a, d such that*

$$\delta(k) \geq \left(1 - \frac{10}{\ell_k \delta(k-1)}\right) \delta(k-1), \qquad k \geq k_0. \tag{3.3.7}$$

Theorem 3.3.6 can only be useful if we know that $\delta(k)$ is much larger than $1/\ell_k$ for k large enough, and thus it is not sufficient to prove Theorem 3.3.2. The next result allows us to have an independent control on $\delta(k)$ which, together with Theorem 3.3.6 implies the desired uniform bound of Theorem 3.3.2.

Theorem 3.3.7. *Assume $SSM(K, a)$. For any $\varepsilon > 0$, there exists a constant $k_0 \in \mathbb{N}$ depending on K, a, d, ε, such that*

$$\delta(k) \geq \ell_k^{-\varepsilon}, \qquad k \geq k_0. \tag{3.3.8}$$

Theorem 3.3.6 and Theorem 3.3.7 are more than sufficient for our purpose. Indeed, using (3.3.8) and (3.3.7), taking for instance $\varepsilon = 1/2$, we see that

$$\ell_k \delta(k-1) \geq \ell_k^{1/2} = (3/2)^{k/2d} \geq 10(6/5)^{k/d}$$

if k is large enough, and therefore

$$\delta(k) \geq \left(1 - (5/6)^{k/d}\right) \delta(k-1) \geq \delta(k_0) \prod_{j=k_0}^{\infty} (1 - (5/6)^{j/d}). \tag{3.3.9}$$

Lemma 3.3.5 and (3.3.9) imply $\inf_{k\in\mathbb{N}} \delta(k) > 0$, which concludes the proof of Theorem 3.3.2 under the assumption $SSM(K, a)$.

3.3.3 Proof of Theorem 3.3.6

We start with a simple decomposition that will be used in the inductive step. Recall that EA and OA stand for the even and odd sites respectively in the region $A \subset V$, and we use the shorthand notation $E = EV$ and $O = OV$ for the whole region V.

Lemma 3.3.8. *For any $A, B \in \mathbb{F}$ such that $V = A \cup B$, for any $f \geq 0$:*

$$\mathrm{Ent}_E f = \mu_E[\mathrm{Ent}_{EA} f + \mathrm{Ent}_{EB}\mu_{EA} f],$$
$$\mathrm{Ent}_O f = \mu_O[\mathrm{Ent}_{OA} f + \mathrm{Ent}_{OB}\mu_{OA} f].$$

Proof. The decomposition in Lemma 2.4.1 shows that

$$\mathrm{Ent}_E f = \mu_E[\mathrm{Ent}_{EA} f] + \mathrm{Ent}_E \mu_{EA} f.$$

Another application of that decomposition shows that

$$\mathrm{Ent}_E \mu_{EA} f = \mu_E[\mathrm{Ent}_{EB}\mu_{EA} f] + \mathrm{Ent}_E \mu_{EB}\mu_{EA} f.$$

However, the product property of μ_E implies that $\mu_{EB}\mu_{EA} f = \mu_E f$, and therefore

$$\mathrm{Ent}_E \mu_{EB}\mu_{EA} f = 0.$$

The same argument applies to the case of odd sites. □

Let us give a sketch of the main steps of the proof before entering the details. Suppose that $V = A \cup B \in \mathbb{F}_k$, and suppose that the assumption of Lemma 3.2.1

is satisfied. Then

$$\mathrm{Ent} f \leq \mu \left[\mathrm{Ent}_A f + \mathrm{Ent}_B \mu_A f \right] + \theta(\varepsilon) \, \mathrm{Ent} f$$

where we use the fact that $\mathrm{Ent}\mu_A f \leq \mathrm{Ent} f$. Now suppose furthermore that $A, B \in \mathbb{F}_{k-1}$. By definition of $\delta(k)$ we then have

$$\begin{aligned} \delta(k-1)\mu[\mathrm{Ent}_A f] &\leq \mu[\mathrm{Ent}_{EA} f + \mathrm{Ent}_{OA} f] \, , \\ \delta(k-1)\mu[\mathrm{Ent}_B(\mu_A f)] &\leq \mu[\mathrm{Ent}_{EB}\mu_A f + \mathrm{Ent}_{OB}\mu_A f]. \end{aligned}$$

Therefore, using Lemma 3.3.8,

$$\begin{aligned} \delta(k-1)\mathrm{Ent} f \leq \mu[\mathrm{Ent}_E f + \mathrm{Ent}_O f] + \theta(\varepsilon)\delta(k-1)\mathrm{Ent} f + \\ + \mu[\mathrm{Ent}_{EB}\mu_A f - \mathrm{Ent}_{EB}\mu_{EA} f + \mathrm{Ent}_{OB}\mu_A f - \mathrm{Ent}_{OB}\mu_{OA} f]. \end{aligned} \tag{3.3.10}$$

Disregarding the second line in (3.3.10) would allow us to obtain a bound of the form

$$\delta(k) \geq (1 - \theta(\varepsilon))\delta(k-1),$$

provided that an arbitrary set $V \in \mathbb{F}_k$ can be decomposed into sets $A, B \in \mathbb{F}_{k-1}$ as above. We remark that if μ were a product over A, B then by convexity one would have

$$\mu[\mathrm{Ent}_{EB}\mu_A f] \leq \mu[\mathrm{Ent}_{EB}\mu_{EA} f], \tag{3.3.11}$$

and the same bound for odd sites. Thus in the product case the second line in (3.3.10) may be neglected and we recover a factorization statement which is contained already in Lemma 3.1.1. In the case we are interested in however one has $A \cap B \neq \emptyset$ and we cannot hope for a bound like (3.3.11). For an illustration of the problem, consider for instance the 1D case, with $V = \{1, \ldots, n\}$, $A = \{1, \ldots, m\}$ and $B = \{m - \ell, \ldots, n\}$ for some integers $0 < \ell < m < n$. Suppose that $m + 1$ is even, and suppose that f only depends on σ_m, the spin at site m. Then, once all odd sites have been frozen, $\mu_{EA} f$ is a constant, and therefore $\mathrm{Ent}_{EB}\mu_{EA} f = 0$. On the other hand, $\mu_A f$ depends on σ_{m+1}, since the conditional expectation μ_A depends non-trivially on σ_{m+1}, and thus we may well have $\mathrm{Ent}_{EB}\mu_A f \neq 0$.

Therefore, the second line of (3.3.10) does produce a nontrivial error term. At this point a fruitful idea from [95] comes to our rescue. Namely, one can average over many possible choices of the decomposition $V = A \cup B$ and hope that the averaging lowers the size of the overall error. This strategy works very well if the error terms have an additive structure, such as in the case of [36]. Here there is no

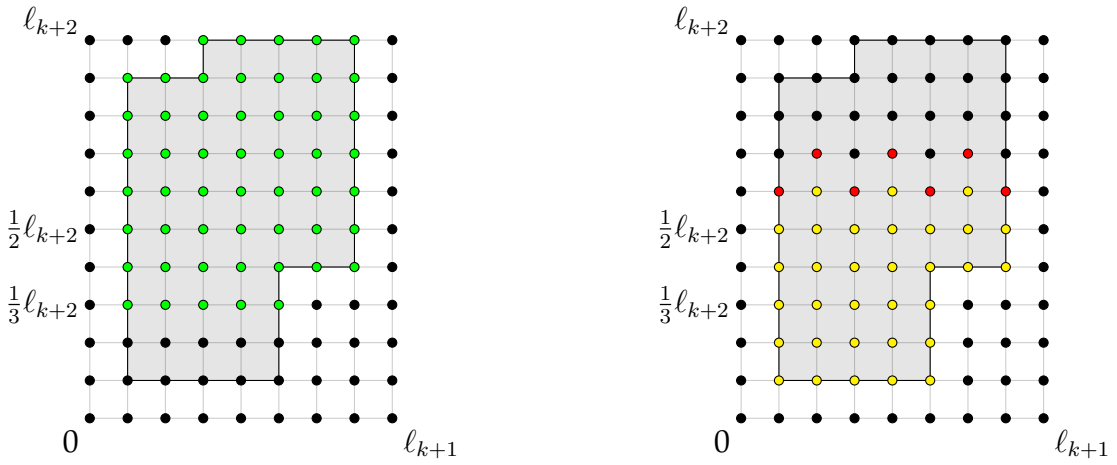

Figure 3.3.1: The gray area is the volume V. Left: B is the set of green vertices. Right: A_1 is the set of yellow vertices, Γ_2 is the set of red vertices, and A_2 is the set of yellow and red vertices together.

simple additive structure to exploit, and we resort to using the martingale-type decompositions from Lemma 2.4.1 to control the average error term by means of the global entropy $\mathrm{Ent} f$, see Lemma 3.3.11. This will be sufficient to obtain the recursive estimate (3.3.7). To implement this argument, we use a slightly different averaging procedure than in [36].

We turn to the actual proof. We start with some geometric considerations, see Figure 3.3.1 for a two-dimensional representation. Set $r := \lfloor \frac{1}{6}\,\ell_{k+d} \rfloor$, and define the rectangular sets

$$
\begin{aligned}
Q &:= [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d-1}] \times [\tfrac{1}{3}\,\ell_{k+d}, \ell_{k+d}] \\
R_i &:= [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d-1}] \times [0, \tfrac{1}{2}\,\ell_{k+d} + i]\,, \quad i = 0, \ldots, r+1.
\end{aligned}
$$

Suppose that $V \subset [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d}]$, and define, for $i = 1, \ldots, r+1$:

$$
B := Q \cap V\,, \quad \text{and} \quad A_i := \begin{cases} (R_i \cap E) \cup (R_{i-1} \cap O) & \text{if } i \text{ is even} \\ (R_i \cap O) \cup (R_{i-1} \cap E) & \text{if } i \text{ is odd} \end{cases}
$$

where, as usual $E = EV$ and $O = OV$ denote the even and the odd sites of V respectively. Define also

$$
\Gamma_i = A_i \setminus A_{i-1}\,, \qquad i = 2, \ldots, r+1.
$$

Lemma 3.3.9. *Suppose that $V \subset [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d}]$, and that $V \notin \mathbb{F}_{k-1}$. Referring to the above setting, for all $i = 1, \ldots r$:*

1. *$V \setminus B \neq \emptyset$ and $V \setminus A_i \neq \emptyset$;*

2. *$d(V \setminus B, V \setminus A_i) \geq \frac{1}{4}\,\ell_k$;*

3. *$B \in \mathbb{F}_{k-1}$ and $A_i \in \mathbb{F}_{k-1}$;*

4. *$\Gamma_{i+1} \subset E$ if i is odd, and $\Gamma_{i+1} \subset O$ if i is even. Moreover A_i and $V \setminus A_{i+1}$ become independent if we condition on the spins in Γ_{i+1}, that is*

$$\mu_V\left(\cdot|\sigma_{\Gamma_{i+1}}\right) = \mu_{V\setminus\Gamma_{i+1}} = \mu_{A_i}\mu_{V\setminus A_{i+1}} = \mu_{V\setminus A_{i+1}}\mu_{A_i}.$$

Proof. 1. Suppose that $V\setminus B$ is empty. Then $V = B$ and therefore, up to translation it is contained in $[0, \ell_{k+1}] \times \cdots \times [0, \frac{2}{3}\ell_{k+d}]$. Since $\frac{2}{3}\ell_{k+d} = \ell_k$ this would imply that up to permutation of the coordinates $V \in [0, \ell_k] \times [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d-1}]$ which violates the assumption $V \notin \mathbb{F}_{k-1}$. The same argument shows that $R_{i-1} \cap V \neq \emptyset$ for all i and $A_i \neq \emptyset$ follows from $A_i \supset R_{i-1} \cap V$.

2. If $x \in V \setminus B$ and $y \in V \setminus A_i$ then $y_d - x_d \geq \frac{1}{2}\ell_{k+d} - \frac{1}{3}\ell_{k+d} = \frac{1}{6}\ell_{k+d} = \frac{1}{4}\ell_k$.

3. The maximal stretch of B along the d-th coordinate is at most $\frac{2}{3}\ell_{k+d} = \ell_k$ and therefore up to translations and permutation of the coordinates $B \in [0, \ell_k] \times [0, \ell_{k+1}] \times \cdots \times [0, \ell_{k+d-1}]$ which says that $B \in \mathbb{F}_{k-1}$. The same argument shows that $A_i \subset R_i \cap V \in \mathbb{F}_{k-1}$ for all i.

4. If $i \geq 1$ is odd, then

$$\begin{aligned}\Gamma_{i+1} &= [(R_{i+1} \cap E) \cup (R_i \cap O)] \setminus [(R_i \cap O) \cup (R_{i-1} \cap E)] \\ &= (R_{i+1} \cap E) \setminus (R_{i-1} \cap E),\end{aligned}$$

and therefore $\Gamma_{i+1} \subset E$. Similarly, one has $\Gamma_{i+1} \subset O$ if i is even. Moreover, any $\mathbb{Z}^d$-path inside V connecting A_i with $V \setminus A_{i+1}$ must go through Γ_{i+1}, and therefore A_i and $V \setminus A_{i+1}$ become independent if we condition on the spins in Γ_{i+1}. □

Lemma 3.3.10. *Let V, B and A_i be as in Lemma 3.3.9. If $SSM(K, a)$ holds, then*

$$\|\mu_B\mu_{A_i}g - \mu g\|_\infty \leq \varepsilon_k \mu(|g|)\,, \qquad \varepsilon_k = 5^d K \ell_k^{d-1} e^{-a\ell_k/4}\,,$$

for all $i = 1, \ldots r$, all functions $g \in L^1(\mu)$, and for all $k \geq k_0 = k_0(K, a, d)$.

Proof. Since i is fixed, for simplicity we write A instead of A_i. Set $h = \mu_A g$. Then h depends only on σ_Δ, where $\Delta = V \setminus A \subset B$. We are going to use (2.2.1) with $\Lambda = B$. Let $\Omega_{B,\tau}$ denote the set of all spin configurations $\eta \in \Omega_{B^c}$ which agree on

the set V^c with the overall boundary condition $\tau \in \Omega_{V^c}$. For any $\eta \in \Omega_{B,\tau}$ one has

$$\begin{aligned}\mu_B^\eta(\mu_A g) - \mu g &= \int \mu_{V,V\setminus B}^\tau(d\eta')\left(\mu_{B,\Delta}^\eta h - \mu_{B,\Delta}^{\eta'} h\right)\\ &= \int \mu_{V,V\setminus B}^\tau(d\eta')\int \mu_{B,\Delta}^{\eta'}(d\sigma)\left(\frac{\psi_{B,\Delta}^\eta(\sigma)}{\psi_{B,\Delta}^{\eta'}(\sigma)} - 1\right)h(\sigma).\end{aligned}$$

Therefore,

$$\|\mu_B\mu_A g - \mu g\|_\infty \le \varepsilon\,\mu(|h|) \le \varepsilon\,\mu(|g|), \tag{3.3.12}$$

where

$$\varepsilon := \sup_{\eta,\eta'\in\Omega_{B,\tau}} \left\| \frac{\psi_{B,\Delta}^\eta}{\psi_{B,\Delta}^{\eta'}} - 1 \right\|_\infty . \tag{3.3.13}$$

Since $\psi_{B,\Delta}^\eta$ depends on η only through the spins in ∂B, the configurations $\eta, \eta' \in \Omega_{B,\tau}$ in (3.3.13) can be assumed to differ only in the set $N_B = (\partial B) \cap (V \setminus B)$. Notice that N_B has at most $(\ell_{k+d-1}+1)^{d-1}$ elements, and that

$$d(N_B, \Delta) \ge d(V\setminus B, V\setminus A) \ge \tfrac{1}{4}\ell_k,$$

by Lemma 3.3.9(2). Therefore, if $\eta(0) = \eta, \dots, \eta(m) = \eta'$, denotes a sequence of configurations interpolating between η and η', such that, for all $j \in \{0,\dots,m-1\}$, $\eta(j)$ and $\eta(j+1)$ differ only at one site $x_j \in N_B$, with $m \le (\ell_{k+d-1}+1)^{d-1}$, we have

$$\frac{\psi_{B,\Delta}^\eta}{\psi_{B,\Delta}^{\eta'}} = \prod_{j=1}^m \frac{\psi_{B,\Delta}^{\eta(j-1)}}{\psi_{B,\Delta}^{\eta(j)}}.$$

The definition of $SSM(K,a)$ implies that

$$\left\| \frac{\psi_{B,\Delta}^{\eta(j-1)}}{\psi_{B,\Delta}^{\eta(j)}} - 1 \right\|_\infty \le \varepsilon_0 := Ke^{-a\ell_k/4}.$$

Expanding the products in (3.3.13), and assuming $m\varepsilon_0 \le 1$, we obtain

$$\varepsilon \le \sum_{\ell=1}^m \binom{m}{\ell}\varepsilon_0^\ell = (1+\varepsilon_0)^m - 1 \le em\varepsilon_0,$$

where we use the inequality $(1+x)^m \le 1 + emx$ for $x > 0$ and $m > 0$ such that $mx \le 1$. Thus, if $k \ge k_0$ for some constant k_0 depending only on K, a, d, we have obtained (3.3.12) with $\varepsilon = K'\ell_k^{d-1}e^{-a\ell_k/4}$, where $K' = (3/2)^{d-1}eK \le 5^dK$. □

Lemma 3.3.11. *Let V, B and A_i, $i = 1, \dots r$, be as in Lemma 3.3.9. Then*

$$\sum_{i=1}^{r} \mu \left[\mathrm{Ent}_{EB} \mu_{A_i} f - \mathrm{Ent}_{EB} \mu_{EA_i} f \right] \leq \mathrm{Ent} f \,,$$
$$\sum_{i=1}^{r} \mu \left[\mathrm{Ent}_{OB} \mu_{A_i} f - \mathrm{Ent}_{OB} \mu_{OA_i} f \right] \leq \mathrm{Ent} f .$$

Proof. We prove the first inequality. The same argument proves the second one, with the role of even and odd sites exchanged. Fix $i \in \{1, \dots, r\}$. Notice that $\mu_{A_i} f = \mu_{A_i} \mu_{EA_i} f$. Let us first observe that if i is even then

$$\mu \left[\mathrm{Ent}_{EB} \mu_{A_i} f - \mathrm{Ent}_{EB} \mu_{EA_i} f \right] \leq 0. \tag{3.3.14}$$

Indeed, in this case $i+1$ is odd and Lemma 3.3.9(4) implies

$$\mu_{EB} \mu_{A_i} f = \mu_{EB} \mu_{A_i} \mu_{EA_i} f = \mu_{A_i} \mu_{EB} \mu_{EA_i} f. \tag{3.3.15}$$

Therefore,

$$\begin{aligned} \mu \left[\mathrm{Ent}_{EB} \mu_{A_i} f \right] &= \mu \left[\mu_{A_i} f \log \left(\mu_{A_i} f / \mu_{EB} \mu_{A_i} f \right) \right] \\ &= \mu \left[\mu_{A_i} \mu_{EA_i} f \log \left(\mu_{A_i} \mu_{EA_i} f / \mu_{A_i} \mu_{EB} \mu_{EA_i} f \right) \right] \\ &= \mu \left[\mu_{EA_i} f \log \left(\mu_{A_i} \mu_{EA_i} f / \mu_{A_i} \mu_{EB} \mu_{EA_i} f \right) \right] \\ &= \mu \left[\mu_{EB} \left(\mu_{EA_i} f \log \left(\mu_{A_i} \mu_{EA_i} f / \mu_{EB} \mu_{A_i} \mu_{EA_i} f \right) \right) \right] \\ &\leq \mu \left[\mu_{EB} \left(\mu_{EA_i} f \log \left(\mu_{EA_i} f / \mu_{EB} \mu_{EA_i} f \right) \right) \right] \\ &= \mu \left[\mathrm{Ent}_{EB} \mu_{EA_i} f \right], \end{aligned} \tag{3.3.16}$$

where the inequality follows from the variational principle (2.3.2). This settles the case when i is even.

Next, suppose that i is odd. Here the commutation relation (3.3.15) does not hold, since the average μ_{A_i} depends on the spins in the even sites $\Gamma_{i+1} \subset B \setminus A_i$. Moreover, (3.3.14) is in general false since if e.g. f depends only on σ_{Γ_i}, then $\mathrm{Ent}_{EB} \mu_{EA_i} f = 0$ while one can have $\mathrm{Ent}_{EB} \mu_{A_i} f > 0$.

Define $g = \mu_{EA_i} f$. From the decomposition in Lemma 2.4.1 we see that

$$\begin{aligned} \mathrm{Ent}_{EB}(\mu_{EA_i} f) &= \mathrm{Ent}_E(g) \\ &= \mathrm{Ent}_E \left(\mu_E(g | \sigma_{i+1}) \right) + \mu_E \left[\mathrm{Ent}_E \left(g | \sigma_{i+1} \right) \right], \end{aligned} \tag{3.3.17}$$

where we use the shorthand notation σ_{i+1} for $\sigma_{\Gamma_{i+1}}$, $\mathrm{Ent}_E \left(g | \sigma_{i+1} \right)$ denotes the entropy of g with respect to the conditional measure $\mu_E(\cdot | \sigma_{i+1}) = \mu_{E \setminus \Gamma_{i+1}}$. Since μ_E

is a product measure,

$$\mathrm{Ent}_E\left(\mu_E(g|\sigma_{i+1})\right) = \mathrm{Ent}_{i+1}\left(\mu_E(g|\sigma_{i+1})\right), \tag{3.3.18}$$

where $\mathrm{Ent}_{i+1} = \mathrm{Ent}_{\Gamma_{i+1}}$ denotes the entropy with respect to the probability measure $\mu_{\Gamma_{i+1}}$. Similarly,

$$\begin{aligned}\mathrm{Ent}_{EB}(\mu_{A_i} f) &= \mathrm{Ent}_E(\mu_{A_i} g) \\ &= \mathrm{Ent}_{i+1}\left(\mu_E(\mu_{A_i} g|\sigma_{i+1})\right) + \mu_E\left[\mathrm{Ent}_E\left(\mu_{A_i} g|\sigma_{i+1}\right)\right].\end{aligned} \tag{3.3.19}$$

Let us show that

$$\mu\left[\mathrm{Ent}_E\left(\mu_{A_i} g|\sigma_{i+1}\right)\right] \le \mu\left[\mathrm{Ent}_E\left(g|\sigma_{i+1}\right)\right]. \tag{3.3.20}$$

Indeed, Lemma 3.3.9(4) implies that

$$\mu_E(\mu_{A_i} g|\sigma_{i+1}) = \mu_{E(V\setminus A_{i+1})}\mu_{A_i} g = \mu_{A_i}\mu_{E(V\setminus A_{i+1})} g = \mu_{A_i}\mu_E(g|\sigma_{i+1}), \tag{3.3.21}$$

where $E(V \setminus A_{i+1})$ are the even sites in $V \setminus A_{i+1}$, and we have used the fact that A_i and $E(V \setminus A_{i+1})$ are conditionally independent given the spins σ_{i+1}. Therefore, reasoning as in (3.3.16):

$$\begin{aligned}\mu\left[\mathrm{Ent}_E\left(\mu_{A_i} g|\sigma_{i+1}\right)\right] &= \mu\left[\mu_{A_i} g \log\left(\mu_{A_i} g/\mu_E(\mu_{A_i} g|\sigma_{i+1})\right)\right] \\ &= \mu\left[\mu_{A_i} g \log\left(\mu_{A_i} g/\mu_{A_i}\mu_E(g|\sigma_{i+1})\right)\right] \\ &= \mu\left[g \log\left(\mu_{A_i} g/\mu_E(\mu_{A_i} g|\sigma_{i+1})\right)\right] \\ &\le \mu\left[g \log\left(g/\mu_E(g|\sigma_{i+1})\right)\right] \\ &= \mu\left[\mathrm{Ent}_E\left(g|\sigma_{i+1}\right)\right].\end{aligned}$$

From (3.3.17)-(3.3.18)-(3.3.19)-(3.3.20) we conclude that, when i is odd:

$$\begin{aligned}&\mu\left[\mathrm{Ent}_{EB}\mu_{A_i} f - \mathrm{Ent}_{EB}\mu_{EA_i} f\right] \\ &\quad\le \mu[\mathrm{Ent}_{i+1}\left(\mu_E(\mu_{A_i} g|\sigma_{i+1})\right)] - \mu[\mathrm{Ent}_{i+1}\left(\mu_E(g|\sigma_{i+1})\right)].\end{aligned} \tag{3.3.22}$$

As in (3.3.21), we may write

$$\mu_E(\mu_{A_i} g|\sigma_{i+1}) = \mu_E(\mu_{A_i} f|\sigma_{i+1}) = \mu_{E(V\setminus A_{i+1})}\mu_{A_i} f\,.$$

Therefore

$$\begin{aligned}\mu[\mathrm{Ent}_{i+1}(\mu_E(\mu_{A_i}g|\sigma_{i+1}))] &= \mu\left[\mathrm{Ent}_{i+1}\left(\mu_{E(V\setminus A_{i+1})}\mu_{A_i}f\right)\right]\\ &\le \mu\left[\mu_{E(V\setminus A_{i+1})}\mathrm{Ent}_{i+1}\mu_{A_i}f\right]\\ &= \mu\left[\mathrm{Ent}_{i+1}\mu_{A_i}f\right],\\ &\le \mu\left[\mathrm{Ent}_{A_{i+1}}\mu_{A_i}f\right],\end{aligned}$$

where the first inequality follows from convexity of entropy and the second from the monotonicity of $A \mapsto \mu[\mathrm{Ent}_A f]$. Neglecting the last term in (3.3.22), we have arrived at

$$\mu\left[\mathrm{Ent}_{EB}\mu_{A_i}f - \mathrm{Ent}_{EB}\mu_{EA_i}f\right] \le \mu\left[\mathrm{Ent}_{A_{i+1}}\mu_{A_i}f\right], \tag{3.3.23}$$

for all i odd. In view of the estimate (3.3.14) we may use the bound (3.3.23) for all i. Therefore, an application of Lemma 2.4.1 shows that

$$\begin{aligned}\sum_{i=1}^{r}\mu\left[\mathrm{Ent}_{EB}\mu_{A_i}f - \mathrm{Ent}_{EB}\mu_{EA_i}f\right] &\le \sum_{i=1}^{r}\mu\left[\mathrm{Ent}_{A_{i+1}}\mu_{A_i}f\right]\\ &= \mu\left[\mathrm{Ent}_{A_{r+1}}\mu_{A_1}f\right] \le \mathrm{Ent}f.\end{aligned}$$

□

We are now able to conclude the proof of Theorem 3.3.6. To prove the recursive bound (3.3.7) we suppose $V \in \mathbb{F}_k \setminus \mathbb{F}_{k-1}$. Then, by translation invariance and by the invariance under coordinate permutation, we may assume that V is as in Lemma 3.3.9. Combining Lemma 3.2.1 with Lemma 3.3.10 we obtain, for each $i = 1, \dots, r$,

$$(1-\theta(\varepsilon_k))\,\mathrm{Ent}f \le \mu\left[\mathrm{Ent}_{A_i}f + \mathrm{Ent}_B\mu_{A_i}f\right].$$

Since $A_i, B \in \mathbb{F}_{k-1}$, by definition of $\delta(k)$ we obtain

$$\begin{aligned}&(1-\theta(\varepsilon_k))\delta(k-1)\,\mathrm{Ent}f \\ &\quad\le \mu\left[\mathrm{Ent}_{EA_i}f + \mathrm{Ent}_{EB}\mu_{A_i}f + \mathrm{Ent}_{OA_i}f + \mathrm{Ent}_{OB}\mu_{A_i}f\right].\end{aligned} \tag{3.3.24}$$

From Lemma 3.3.8 we find that the right hand side of (3.3.24) equals

$$\begin{aligned}&\mu\left[\mathrm{Ent}_E f + \mathrm{Ent}_O f\right] + \\ &\quad+ \mu\left[\mathrm{Ent}_{EB}\mu_{A_i}f - \mathrm{Ent}_{EB}\mu_{EA_i}f\right] + \mu\left[\mathrm{Ent}_{OB}\mu_{A_i}f - \mathrm{Ent}_{OB}\mu_{OA_i}f\right].\end{aligned} \tag{3.3.25}$$

Averaging over i in (3.3.25) and using Lemma 3.3.11,

$$(1-\theta(\varepsilon_k))\delta(k-1)\,\mathrm{Ent} f \le \mu\,[\mathrm{Ent}_E f + \mathrm{Ent}_O f] + \frac{2}{r}\,\mathrm{Ent} f.$$

In conclusion, $\delta(k) \ge (1-\theta(\varepsilon_k))\delta(k-1) - \frac{2}{r}$, or equivalently

$$\delta(k) \ge \left(1 - \theta(\varepsilon_k) - \frac{2}{r\delta(k-1)}\right)\delta(k-1).$$

Since $r \sim \frac{1}{4}\ell_k$ and $\delta(k-1) \le 1$, it follows that $\frac{1}{r\delta(k-1)} \gg \theta(\varepsilon_k)$ for all k large enough, and therefore

$$\delta(k) \ge \left(1 - \frac{10}{\delta(k-1)\ell_k}\right)\delta(k-1),$$

for all $k \ge k_0(K,a,d)$.

3.3.4 Proof of Theorem 3.3.7

The idea is to divide the set V into two sets $A = \cup_i A_i$, $B = \cup_i B_i$ each being the union of a large number of well separated subsets, to use the factorization from Lemma 3.2.1 to reduce the problem in the set V to the problem in either A or B, and then finally to use the Lemma 3.1.1 to tensorize within A and within B, which allows us to reduce the problem to a single region A_i or B_i only.

Fix a large integer $b > 1$, define $u_k = b^{k/d}$, and call $\mathbb{G}_k$ the set of all subsets $V \subset \mathbb{Z}^d$ which up to translations and permutation of the coordinates are included in the rectangle $[0, u_{k+1}] \times \cdots \times [0, u_{k+d}]$. We partition the interval $I = [0, u_{k+d}]$ into $2b$ consecutive non-overlapping intervals $I_1, \ldots, I_{2b}$ such that I_j have length $t_k := \frac{1}{2b}u_{k+d}$, that is

$$I_j = [(j-1)t_k, jt_k], \qquad j = 1, \ldots, 2b.$$

Define also the enlarged intervals $\bar{I}_j = \{s \in I : d(s, I_j) \le t_k/4\}$, and consider the collections of intervals

$$\Delta_A = \bigcup_{j=1}^{2b} \bar{I}_j\, \mathbf{1}_{j\,\text{odd}}, \qquad \Delta_B = \bigcup_{j=1}^{2b} \bar{I}_j\, \mathbf{1}_{j\,\text{even}}.$$

We remark that both Δ_A and Δ_B are collections of non-overlapping intervals, with

$$d(\bar{I}_{2j-1}, \bar{I}_{2i-1}) \ge \frac{1}{2}\,t_k\,, \qquad d(\bar{I}_{2j}, \bar{I}_{2i}) \ge \frac{1}{2}\,t_k$$

for all $i \ne j$. On the other hand, $\Delta_A \cap \Delta_B \ne \emptyset$. We define the rectangular sets in

$\mathbb{R}^d$:

$$Q_i := [0, u_{k+1}] \times \cdots \times [0, u_{k+d-1}] \times \bar{I}_j \,, \qquad j = 1, \ldots, 2b,$$

and define the $\mathbb{Z}^d$ subsets

$$A_i := Q_{2i-1} \cap V \,, \quad B_i = Q_{2i} \cap V \,, \qquad i = 1, \ldots, b.$$
$$A = \bigcup_{i=1}^{b} A_i \,, \qquad B = \bigcup_{i=1}^{b} B_i \,.$$

We refer to Figure 3.3.2 for a two-dimensional representation.

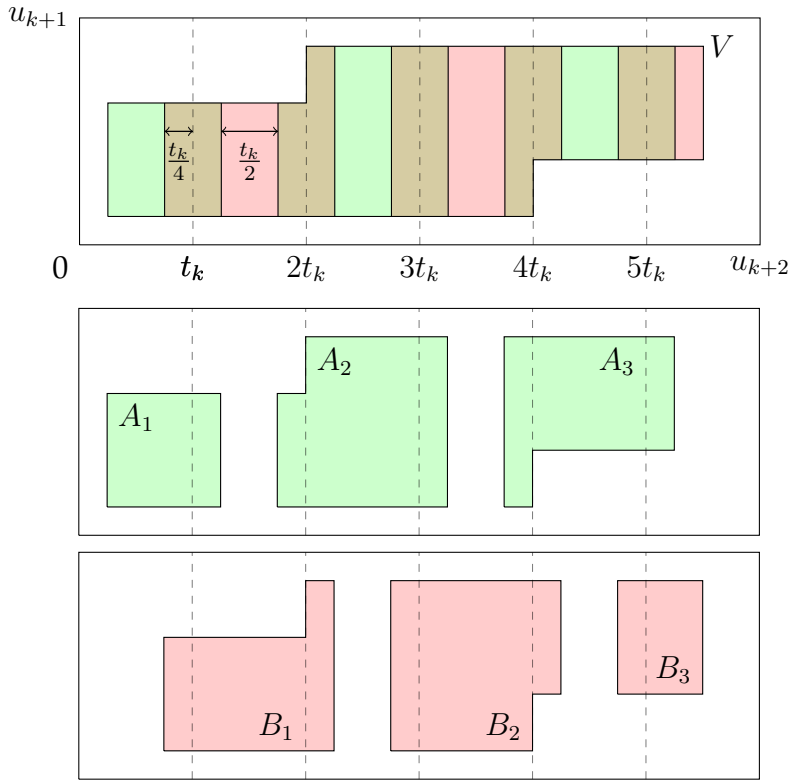

Figure 3.3.2: An example of $A = \bigcup_i A_i$ (green blocks) and $B = \bigcup_i B_i$ (red blocks) for a given region V in the rectangle $[0, u_{k+1}] \times [0, u_{k+2}]$.

We observe that $A_i \in \mathbb{G}_{k-1}$ and $B_i \in \mathbb{G}_{k-1}$ for all $i = 1, \ldots, b$. Indeed, the stretch of A_i along the d-th coordinate is at most $t_k + 2t_k/4 \leq 2t_k \leq u_k$ which together with $u_{k,i} = u_{k-1,i+1}$, $i = 1, \ldots, d-1$, implies that $A_i \in \mathbb{G}_{k-1}$. The same applies to B_i. Observe that with these definitions one has the product property

$$\mu_A = \otimes_{i=1}^{b} \mu_{A_i} \,, \qquad \mu_B = \otimes_{i=1}^{b} \mu_{B_i}. \tag{3.3.26}$$

Moreover, the geometric construction shows that

$$d(V \setminus A, V \setminus B) \geq \frac{1}{2} t_k.$$

Thus, a repetition of the argument in Lemma 3.3.10 shows that the assumption of Lemma 3.2.1 is satisfied with ε given by

$$\varepsilon_k = O\left(u_k^d \, e^{-a u_k/2}\right).$$

Therefore, by Lemma 3.2.1,

$$\operatorname{Ent} f \leq \mu\left[\operatorname{Ent}_A f + \operatorname{Ent}_B f\right] + \theta(\varepsilon_k) \operatorname{Ent} f. \tag{3.3.27}$$

Next, let $\varrho(k)$ be defined as the largest constant $\varrho > 0$ such that the inequality

$$\varrho \operatorname{Ent}_V^\tau f \leq \mu_V^\tau \left[\operatorname{Ent}_E f + \operatorname{Ent}_O f\right]$$

holds for all $V \in \mathbb{G}_k$, $\tau \in \Omega_{V^c}$, and all $f \geq 0$. The key observation is that thanks to the product property (3.3.26), and using the fact that $A_i \in \mathbb{G}_{k-1}$ for all i, Lemma 3.1.1 allows us to estimate

$$\varrho(k-1)\mu\left[\operatorname{Ent}_A f\right] \leq \mu\left[\operatorname{Ent}_{EA} f + \operatorname{Ent}_{OA} f\right].$$

Similarly,

$$\varrho(k-1)\mu\left[\operatorname{Ent}_B f\right] \leq \mu\left[\operatorname{Ent}_{EB} f + \operatorname{Ent}_{OB} f\right].$$

Thus, (3.3.27) implies

$$\begin{aligned}\varrho(k-1)(1-\theta(\varepsilon_k))\operatorname{Ent} f &\leq \mu\left[\operatorname{Ent}_{EA} f + \operatorname{Ent}_{OA} f\right] + \mu\left[\operatorname{Ent}_{EB} f + \operatorname{Ent}_{OB} f\right] \\ &\leq 2\mu\left[\operatorname{Ent}_E f + \operatorname{Ent}_O f\right],\end{aligned}$$

where we use the monotonicity of $\Lambda \mapsto \mu\left[\operatorname{Ent}_\Lambda f\right]$. Estimating $1 - \theta(\varepsilon_k) \geq 1/2$ we have proved that

$$\varrho(k) \geq \frac{1}{4}\varrho(k-1).$$

Iterating, we conclude $\varrho(k) \geq 4^{-k}\varrho(k_0)$. To finish the proof, observe that $(3/2)^k = b^{k\varepsilon}$ where $\varepsilon = \log(3/2)/\log(b)$, which can be made small by taking b large. There-

fore,

$$\delta(k) \geq \varrho(\lfloor k\varepsilon \rfloor + 1) \geq 4^{-k\varepsilon - 1}\varrho(k_0) \geq c_0 4^{-k\varepsilon} = c_0 \ell_k^{-\varepsilon'},$$

where c_0 is a constant depending on K, a, d, b, while $\varepsilon' = d\log(4)/\log(b)$ can be as small as we wish provided b is suitably large. This ends the proof of Theorem 3.3.7.

Remark 3.3.12. We point out that the argument given in the proof of Theorem 3.3.7 can be improved if one replaces the parameter t_k which is linear in u_k by $t'_k = C_1 \log(u_k)$, with C_1 a suitably large constant. Since t'_k is logarithmic in u_k, one can modify the recursion to obtain a bound of the form $\delta(k) \geq \delta(C_2 \log(k))/C_2$ for some new constant C_2, which provides a much better lower bound on $\delta(k)$ than the one stated in Theorem 3.3.7. However, without the companion recursive estimate from Theorem 3.3.6, this argument alone would not provide the uniform estimate $\inf_k \delta(k) > 0$.

3.3.5 Proof of Theorem 3.3.2 assuming $SSM_L(K, a)$

Theorem 3.3.6 and Theorem 3.3.7 allowed us to establish Theorem 3.3.2 under the assumption $SSM(K, a)$. We now prove it assuming only $SSM_L(K, a)$. To this end we observe that any set $V \in \mathbb{F}^{(L)}$ is uniquely identified by the set $V' \in \mathbb{F}$ such that

$$V = \bigcup_{y \in V'} Q_L(y). \tag{3.3.28}$$

A careful check of the previous proofs then shows that if we work on the rescaled lattice, that is we replace vertices x with blocks $Q_L(x)$, then we may repeat all steps in Theorem 3.3.6 and Theorem 3.3.7 to obtain the following coarse-grained version of Theorem 3.3.2 assuming only $SSM_L(K, a)$: for any $V \in \mathbb{F}^{(L)}$, for all $f \geq 0$,

$$\mathrm{Ent} f \leq C\, \mu \left[\mathrm{Ent}_{E_L} f + \mathrm{Ent}_{O_L} f \right], \tag{3.3.29}$$

where, if V is given by (3.3.28), then $E_L = \cup_{x \in EV'} Q_L(x)$, and $O_L = \cup_{x \in OV'} Q_L(x)$.

Consider now a single cube $Q_L(x)$. By Lemma 3.3.5 we know that

$$\mathrm{Ent}_{Q_L(x)} f \leq C_1\, \mu_{Q_L(x)} \left[\mathrm{Ent}_{EQ_{L(x)}} f + \mathrm{Ent}_{OQ_L(x)} f \right],$$

for some constant $C_1 = C_1(L)$. Observe that by construction $d(Q_L(x), Q_L(y)) > 1$ for all $x, y \in EV'$. Similarly, $d(Q_L(x), Q_L(y)) > 1$ for all $x, y \in OV'$. Therefore,

Lemma 3.1.1 implies

$$\begin{aligned}\mathrm{Ent}_{E_L} f &\leq C_1\, \mu_{E_L}\left[\mathrm{Ent}_{EE_L} f + \mathrm{Ent}_{OE_L} f\right] \\ \mathrm{Ent}_{O_L} f &\leq C_1\, \mu_{O_L}\left[\mathrm{Ent}_{EO_L} f + \mathrm{Ent}_{OE_L} f\right],\end{aligned}$$

where EE_L denotes the even sites in E_L, EO_L the even sites in O_L, and so on. Plugging these estimates in (3.3.29) and using the monotonicity of $A \mapsto \mu[\mathrm{Ent}_A f]$ one arrives at

$$\mathrm{Ent} f \leq D\, \mu\left[\mathrm{Ent}_E f + \mathrm{Ent}_O f\right],$$

with $D = 2C \times C_1$. This ends the proof of Theorem 3.3.2.

Chapter 4

Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$

4.1 Introduction and main results

In this chapter we consider the Swendsen-Wang dynamics, and we will prove Theorems 1.3.3 and 1.3.5 together with some additional results that we state later. The notion of SSM that we use in this chapter is given in definition 2.2.3. Our main analytic tool establishes that, on any bounded degree bipartite graph, the even/odd factorization is sufficient to ensure $O(\log n)$ mixing time for the SW dynamics. Note that SW is very far from a block dynamics, in that the configurations of multiple, dynamically changing clusters of spins are updated simultaneously in each step.

Theorem 4.1.1. *For all constant Δ, for any bipartite graph of maximum degree Δ, if the Gibbs distribution satisfies approximate even/odd factorization with $C = O(1)$ then the mixing time of the Swendsen-Wang dynamics is $O(\log n)$.*

We remark that Theorem 4.1.1 holds for arbitrary boundary conditions (or pinnings of vertices) of the bipartite graph, and thus the $O(\log n)$ bound for the mixing time of the SW dynamics in Theorem 1.3.3 for the lattice $\mathbb{Z}^d$ follows immediately from this Theorem and Theorem 3.3.2.

The main technical step in the proof of Theorem 4.1.1 is to show that even/odd factorization implies a novel *spin/edge factorization* of entropy (see Definition (4.1.2) below), which is tailored to the SW dynamics so that it implies $O(\log n)$ mixing fairly directly.

4.1.1 The spin/edge factorization

This new entropy factorization is based on the joint probability space on spins and edges introduced by Edwards and Sokal [56], that underlies the SW dynamics. In this chapter we will denote the set of edges of the underlying graph as $\mathbb{E}$, to distinguish it from the set E of the even vertices. Let $\Omega_{\text{J}} = \Omega \times \{0,1\}^{\mathbb{E}}$ be the set of joint configurations (σ, A) consisting of a spin assignment to the vertices $\sigma \in \Omega$ and a subset of edges $A \subseteq \mathbb{E}$, where recall that $\mathbb{E}$ is the set of edges with both endpoints in V. The Edwards-Sokal distribution on G with parameters $p \in [0,1]$ and $q \in \mathbb{N}$, and free boundary condition, is the probability measure on Ω_{J} given by

$$\nu(\sigma, A) := \frac{1}{Z_{\text{J}}}\, p^{|A|}(1-p)^{|\mathbb{E}|-|A|}\mathbf{1}(\sigma \sim A), \tag{4.1.1}$$

where $\sigma \sim A$ means that $A \subseteq M(\sigma)$ (i.e., that 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 $Z = Z_{\text{J}}$; the "edge marginal" of ν corresponds to the well-known random-cluster measure; see [62, 68]. The SW dynamics alternates between spin configurations and joint spin/edge configurations in a manner consistent with (4.1.1).

We note that a boundary condition on the joint space allows fixing the state of *both* spins and edges and thus may introduce more complex dependencies. While our results in the joint space are stated here only for the free boundary condition, they actually extend to any *spin-only* boundary condition. By a "spin-only" boundary condition we mean any boundary condition that fixes the spins of a subset of vertices, and fixes no values for the edges. In fact, in $\mathbb{Z}^d$, we can handle a slightly more general class of boundary conditions we call *admissible* (see Definition 4.3.1) which will be useful for proving Theorem 1.3.5 and our results for random-cluster dynamics.

Our entropy factorization for the SW dynamics is defined as follows.

Definition 4.1.2. We say that *approximate spin/edge factorization* with constant C holds if for all functions $f : \Omega_{\text{J}} \mapsto \mathbb{R}_+$,

$$\text{Ent}_\nu(f) \leq C\ \left(\nu\left[\text{Ent}_\nu(f \,|\, \sigma)\right] + \nu[\text{Ent}_\nu(f \,|\, A)]\right). \tag{4.1.2}$$

Let us explain the terms in (4.1.2) in more detail. We write $\nu(\cdot \,|\, \sigma)$ for the probability obtained from ν by conditioning the on whole spin configuration being equal to a given $\sigma \in \Omega$ and $\nu(\cdot \,|\, A)$ for the probability obtained from ν by conditioning on the whole edge configuration being equal to a given $A \subseteq \mathbb{E}$. With this notation, $\text{Ent}_\nu(f \,|\, \sigma)$ and $\text{Ent}_\nu(f \,|\, A)$ denote the entropy of f with respect to the conditional measures $\nu(\cdot \,|\, \sigma)$ and $\nu(\cdot \,|\, A)$, respectively. Therefore, taking their

expectation with respect to ν one obtains

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma)\right] = \nu[f \log(f/\nu[f \mid \sigma])]\,,$$
$$\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] = \nu[f \log(f/\nu[f \mid A])].$$

The main technical ingredient in proving Theorem 4.1.1 is the following "comparison lemma" for entropy factorization.

Lemma 4.1.3. *For the Potts model at inverse temperature β on any bipartite graph of maximum degree Δ, approximate even/odd factorization with constant C implies approximate spin/edge factorization with constant $C' = C'(C, \Delta, q, \beta)$.*

To complete the proof of Theorem 4.1.1, we show that the spin/edge factorization in (4.1.2) implies relative entropy decay for the SW dynamics, which implies the modified log-Sobolev inequality which finally implies a bound of $O(\log n)$ on the mixing time of the SW dynamics, see Lemma 2.3.3.

Lemma 4.1.4. *For the Potts model on* any *n-vertex graph, approximate spin/edge factorization with constant $C = O(1)$ implies that inequality* (2.3.4) *for SW holds with $\delta = 1/C$ and hence $T_{\mathrm{mix}}(SW) = O(\log n)$.*

4.1.2 Further results

Our new entropy factorization framework leads to several additional algorithmic results on $\mathbb{Z}^d$ that hold under the condition of SSM, which we briefly summarize here. First, we prove optimal $O(\log n)$ mixing time for *alternating systematic scan dynamics*, a natural non-local dynamics in which even and odd sides of the bipartition are updated on alternate steps. Systematic scan dynamics, in which updates are performed in a deterministic rather than random sequence, are widely used in practice but are non-reversible and typically much harder to analyze. Second, we are able to show that various versions of the SW dynamics on the *joint spin/edge space* mix in $O(\log n)$ time, as does the SW dynamics for the *random-cluster model*. Finally, we show that a natural local Glauber dynamics in the joint space has optimal mixing time $\Theta(n \log n)$. Formal statements of all these results will be given later.

Remark 4.1.5. For definiteness, we have stated all of our results for n-vertex d-dimensional cubes but they extend to more general regions of $\mathbb{Z}^d$. In particular, we can consider regions which are the union of disjoint translates of a given large enough cube, and thus we can consider definition 2.2.2 of spatial mixing.

We start with Section 4.2, where we prove Lemma 4.1.4 showing that the spin/edge entropy factorization implies $O(\log n)$ mixing for the SW dynamics.

We then prove Lemma 4.1.3 relating even/odd factorization to spin/edge factorization in Section 4.3, and then combine Lemmas 4.1.3 and 4.1.4 to establish our main technical tool (Theorem 4.1.1) and our main algorithmic result (Theorem 1.3.3). Our lower bound on the mixing time is proved in Section 4.4. We discuss further applications in the remaining sections. Section 4.5 proves entropy decay for non-local and local dynamics in the joint space, and Section 4.6 discusses the alternating scan dynamics. Finally, we address the random-cluster dynamics in Section 4.7, concluding with a proof of Theorem 1.3.5.

4.2 Spin/edge factorization implies fast mixing: proof of Lemma 4.1.4

As mentioned in the introduction, the proof of our main new analytic tool (Theorem 4.1.1) has two components. We show that approximate even/odd factorization implies spin/edge factorization (Lemma 4.1.3), and then that spin/edge factorization implies $O(\log n)$ mixing for the SW dynamics (Lemma 4.1.4). In this section, we provide the proof of the latter result, whereas Lemma 4.1.3 is proved in the subsequent section.

Before starting our analysis, we need the two following identities:

$$\mathrm{Ent}_\nu(f) = \mathrm{Ent}_\nu(\nu[f \,|\, A]) + \nu[\mathrm{Ent}_\nu(f \,|\, A)]; \tag{4.2.1}$$

$$\mathrm{Ent}_\nu(f) = \mathrm{Ent}_\nu(\nu[f \,|\, \sigma]) + \nu[\mathrm{Ent}_\nu(f \,|\, \sigma)]. \tag{4.2.2}$$

Both statements follow from the general decomposition

$$\mathrm{Ent}_\pi(f) = \mathrm{Ent}_\pi(\pi[f \,|\, \mathcal{F}]) + \pi[\mathrm{Ent}_\pi(f \,|\, \mathcal{F})],$$

valid for any distribution π, and any sub σ-algebra $\mathcal{F}$, which follows by adding and subtracting the term $\pi(f \log \pi[f \,|\, \mathcal{F}])$ to $\mathrm{Ent}_\pi(f)$.

Proof of Lemma 4.1.4. We show that the spin/edge factorization with constant C implies that for all functions $f \geq 0$ with $\mu[f] = 1$, one has

$$\mathrm{Ent}_\mu(P_{\mathrm{SW}} f) \leq (1 - \delta)\mathrm{Ent}_\mu(f), \tag{4.2.3}$$

with $\delta = 1/C$. Since the SW dynamics is reversible with respect to μ, we have $P_{\mathrm{SW}} = P^*_{\mathrm{SW}}$, and the desired mixing time bound follows from Lemma 2.3.3 and Remark 2.3.2.

The transition matrix of the SW dynamics satisfies

$$P_{\mathrm{sw}}(\sigma,\tau) = \sum_{A\subseteq M(\sigma)} \nu(A\,|\,\sigma)\nu(\tau\,|\,A),$$

where we recall that $M(\sigma)$ is the set of monochromatic edges in σ. Hence,

$$P_{\mathrm{sw}}f(\sigma) = \sum_{\tau\in\Omega} P_{\mathrm{sw}}(\sigma,\tau)f(\tau) = \sum_{\tau\in\Omega}\sum_{A\subseteq M(\sigma)} \nu(A\,|\,\sigma)\nu(\tau\,|\,A)\hat{f}(\tau,A),$$

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

$$P_{\mathrm{sw}}f(\sigma) = \sum_{A\subseteq M(\sigma)} \nu(A\,|\,\sigma)\nu[\hat{f}\mid A] = \nu[\nu[\hat{f}\,|\,A]\,|\,\sigma] = \nu[g\,|\,\sigma],$$

where for ease of notation we set $g:=\nu[\hat{f}\,|\,A]$. Since $\mu[f]=1$, we have $\mu[P_{\mathrm{sw}}f]=1$ and

$$\mathrm{Ent}_\mu(P_{\mathrm{sw}}f) = \mu[(P_{\mathrm{sw}}f)\log(P_{\mathrm{sw}}f)] = \mu\left[\nu\left[g\,|\,\sigma\right]\log(\nu\left[g\,|\,\sigma\right])\right].$$

The convexity of the function $x\cdot\log x$ and Jensen's inequality imply

$$\nu\left[g\,|\,\sigma\right]\log(\nu\left[g\,|\,\sigma\right]) \le \nu\left[g\log g\mid\sigma\right],$$

and then, since $\nu[g]=\nu[\hat{f}]=\mu[f]=1$, we have

$$\mathrm{Ent}_\mu(P_{\mathrm{sw}}f) \le \mu\left[\nu\left[g\log g\mid\sigma\right]\right] = \nu\left[g\log g\right] = \mathrm{Ent}_\nu(g). \tag{4.2.4}$$

For any function $h:\Omega_{\mathrm{J}}\mapsto\mathbb{R}_+$, we have by (4.2.1) that $\mathrm{Ent}_\nu(h)=\mathrm{Ent}_\nu(\nu[h|A])+\nu[\mathrm{Ent}_\nu(h|A)]$. Hence,

$$\mathrm{Ent}_\nu(\hat{f}) = \mathrm{Ent}_\nu(g) + \nu[\mathrm{Ent}_\nu(\hat{f}|A)],$$

which by (4.2.4) gives $\mathrm{Ent}_\mu(P_{\mathrm{sw}}f)\le\mathrm{Ent}_\nu(\hat{f})-\nu[\mathrm{Ent}_\nu(\hat{f}|A)]$. The function $\hat{f}$ depends on σ only, so $\mathrm{Ent}_\nu(\hat{f}\,|\,\sigma)=0$. Therefore,

$$\mathrm{Ent}_\mu(P_{\mathrm{sw}}f) \le \mathrm{Ent}_\nu(\hat{f}) - \nu\left[\mathrm{Ent}_\nu(\hat{f}\,|\,A) + \mathrm{Ent}_\nu(\hat{f}\,|\,\sigma)\right].$$

The assumed spin/edge factorization (4.1.2) then implies that $\mathrm{Ent}_\mu(P_{\mathrm{sw}}f)\le(1-\delta)\mathrm{Ent}_\nu(\hat{f})$, with $\delta=1/C$. Inequality (4.2.3) follows from the fact that $\mathrm{Ent}_\nu(\hat{f})=$

$\mathrm{Ent}_\mu(f)$. □

Remark 4.2.1. We do not assume anything about the underlying graph in the previous proof, so Lemma 4.1.4 holds for any graph G. In addition, our proof as stated applies to the Potts measure μ obtained as the marginal on spins of the joint measure ν. If ν is as in (4.1.1), this yields only the Potts measure on V with the *free* boundary condition. However, the proof extends to the Potts measure with *any* boundary condition (or pinning of vertices) by choosing a spin-only boundary condition for ν. In particular, Theorem 1.3.3 holds for arbitrary boundary conditions, as stated in the introduction. For the special case when G is a cube of $\mathbb{Z}^d$, we allow a slightly more general class of boundary conditions, involving both spin and edges, which we call *admissible*; see Definition 4.3.1 and the examples immediately following it.

Remark 4.2.2. The entropy contraction established in (4.2.3) implies a modified log-Sobolev inequality, and can be viewed as a discrete time version of it, as pointed out in section 2.3. The classical log-Sobolev constant, however, is not tight for the SW dynamics. Indeed, the remark in [95, Section 3.7] shows a test function f such that $\mathrm{Var}_\mu(\sqrt{f})/\mathrm{Ent}_\mu(f) = O(n^{-1})$. Since $\mathcal{D}_{P_{\mathrm{SW}}}(\sqrt{f},\sqrt{f}) = \nu[\mathrm{Var}(\sqrt{f} \mid A)]$, it follows from monotonicity of variance functional that $\mathcal{D}_{P_{\mathrm{SW}}}(\sqrt{f},\sqrt{f}) \leq \mathrm{Var}_\mu(\sqrt{f})$ and so $\frac{\mathcal{D}_{P_{\mathrm{SW}}}(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\mu(f)} = O(n^{-1})$ for this function.

4.3 Factorization of entropy in the joint space

In this section, we prove our main technical result, Lemma 4.1.3, which states that approximate even/odd factorization implies approximate spin/edge factorization for the Potts measure on bipartite graphs. For clarity of notation, and to simplify the proofs, we will restrict attention to n-vertex cubes in $\mathbb{Z}^d$, but it should be clear that everything extends to arbitrary bipartite graphs of constant degree with any spin-only boundary condition. In addition, on $\mathbb{Z}^d$ we are able to extend our results to a more general class of boundary conditions in the joint space, involving both edges and vertices, that we call *admissible*.

Admissible boundary conditions. Let ∂V be the set of vertices of V with a neighbor in $\mathbb{Z}^d \setminus V$. Let $\partial\mathbb{E}$ denote the set of edges in $\mathbb{E}$ with at least one endpoint in ∂V. (Recall that $\mathbb{E}$ is the set of edges with both endpoints in V.) We consider boundary conditions for the joint space on subsets $V_0 \subseteq \partial V$ and $\mathbb{E}_0 \subseteq \partial\mathbb{E}$. Specifically, we let $\psi : V_0 \mapsto [q]$ and $\varphi : \mathbb{E}_0 \mapsto \{0,1\}$ and define

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

where $\sigma \sim A$ means that $A \subseteq M(\sigma)$, $\sigma \sim \psi$ that σ and ψ agree on the spins in V_0, and $A \sim \varphi$ that A and φ agree on the edges in $\mathbb{E}_0$. As usual, $Z^{\psi,\varphi}$ is the corresponding partition function.

Definition 4.3.1. We call the boundary condition *admissible* if $\mathbb{E}_0 \subset \{\{u, v\} \in \partial\mathbb{E} : u \in V_0\}$; that is, if all edges in $\mathbb{E}_0$ have at least one endpoint in V_0.

Notice that the free boundary condition ($V_0 = \emptyset$ and $\mathbb{E}_0 = \emptyset$) is admissible, and all spin-only boundary conditions ($V_0 \subset \partial V$ and $\mathbb{E}_0 = \emptyset$) are also admissible. In this case, the marginal on spins is just the Potts measure with ψ as the boundary condition on ∂V with $U^\psi = V_0$. For some additional examples of admissible boundary conditions see Section 4.7 and Figure 4.7.1; in particular, (4.7.2) captures the effects an admissible boundary condition may have on the random-cluster marginal.

The main motivation for introducing the notion of admissible boundary conditions is that it guarantees that the spin marginal of $\nu^{\psi,\varphi}$ has the desired exponential decay of correlations if the parameters q and β are such that SSM holds. We shall see that all of our results concerning the joint measure and its dynamics on $\mathbb{Z}^d$ extend to the more general class of admissible boundary conditions. We can therefore restate Lemma 4.1.3 from the introduction for the special case of $\mathbb{Z}^d$ allowing arbitrary admissible boundary conditions.

Lemma 4.3.2. *Let $\nu := \nu^{\psi,\varphi}$ be the joint distribution with an admissible boundary condition (ψ, φ). Approximate even/odd factorization with constant C of the spin marginal of ν implies that approximate spin/edge factorization holds with constant $C' = C'(C, d, q, \beta)$.*

For simplicity, we will continue to write ν for the joint measure $\nu^{\psi,\varphi}$ and μ for its marginal on spins. We shall see that our proofs in this section are largely oblivious to the boundary condition or the geometry of $\mathbb{Z}^d$ (in fact, we only require the underlying graph to be bipartite). We also remark that, while we could allow a slightly more general family of boundary conditions than the admissible ones, some limitations are needed. For instance, arbitrary edge boundary conditions may cause long-range dependencies; see, e.g., [16, 15]. We proceed next with the proof of Lemma 4.3.2.

4.3.1 Proof of Lemma 4.3.2

Overview. The following high level observations might be of help before entering the technical details of the proof. First, notice that the conclusion in the theorem would trivially hold true with constant $C = 1$ if ν were a product measure with respect to the two sets of variables (σ, A). This is a consequence of standard factorization properties of product measures. Thus, the minimal constant C for which

that statement holds is a measure of the "cost" for "separating" the two sets of variables.

When the dependencies between the two sets of variables are very weak, a factorization statement could be obtained as in [36]. However, in our case the dependencies are not weak, since the spin variables interact locally with the edge variable in a strong way. For instance, the presence of the edge xy in A forces deterministically the condition $\sigma_x = \sigma_y$. Thus, the fact that our statement holds with a constant C independent of n is highly nontrivial.

On the other hand, for every $x \in V$ one can separate *locally* the two variables (σ_x, A_x), where A_x denotes the set of edge variables for edges incident to x, by paying a finite cost C; this is the content of Lemma 4.3.4 below. We can then lift this local factorization to a global factorization statement for the conditional measure $\nu(\cdot \mid \sigma_E)$, respectively $\nu(\cdot \mid \sigma_O)$, obtained by conditioning on the spin variables of all even vertices $E \subset V$, respectively of all odd vertices $O \subset V$. This is the content of Lemma 4.3.4.

Lemma 4.3.4 is the heart of the proof and relies crucially on the fact that $\nu(\cdot \mid \sigma_E)$ is a product measure with respect to $\{(\sigma_x, A_x), x \in O\}$, and $\nu(\cdot \mid \sigma_O)$ is a product measure with respect to $\{(\sigma_x, A_x), x \in E\}$. Thus, we reduce the problem of separating the spin/edge variables (σ, A) to the problem of separating the even/odd spin variables (σ_E, σ_O) for the joint distribution ν. We then conclude by showing that even/odd factorization for the Potts measure μ implies the even/odd factorization for ν. This is the content of Lemma 4.3.5.

We now turn to the actual proof. Let $\nu(\cdot \,|\, \sigma_E, A)$ denote the measure ν conditioned on $\sigma_E = \{\sigma_v,\, v \in E\}$ and $A \subseteq \mathbb{E}$. Similarly, $\nu(\cdot \,|\, \sigma_O, A)$ denotes the measure ν conditioned on $\sigma_O = \{\sigma_v,\, v \in O\}$ and A. We use $\mathrm{Ent}_\nu(f \,|\, \sigma_E, A)$ and $\mathrm{Ent}_\nu(f \,|\, \sigma_O, A)$ to denote the corresponding conditional entropies and we write $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_E, A)\right]$, $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_O, A)\right]$ for their expectations with respect to ν. The next lemma shows that conditioning on the spin configuration of the even or the odd sub-lattice can only decrease the entropy of a function with respect to $\nu(\cdot \mid A)$.

Lemma 4.3.3. *For all functions $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$ we have*

$$\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] \geq \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E, A)\right]; \;\; \textit{and}$$
$$\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] \geq \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_O, A)\right].$$

Proof. We can write

$$\begin{aligned}
\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] &= \nu\left[f \log\left(\frac{f}{\nu[f \mid A]}\right)\right] \\
&= \nu\left[f \log\left(\frac{f}{\nu[f \mid \sigma_E, A]}\right)\right] + \nu\left[f \log\left(\frac{\nu[f \mid \sigma_E, A]}{\nu[f \mid A]}\right)\right] \\
&= \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E, A)\right] + \nu\left[\nu[f \mid \sigma_E, A] \log\left(\frac{\nu[f \mid \sigma_E, A]}{\nu[f \mid A]}\right)\right] \\
&= \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E, A)\right] + \nu\left[\mathrm{Ent}_\nu(\nu[f \mid \sigma_E, A] \mid A)\right] \\
&\geq \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E, A)\right].
\end{aligned}$$

The same argument applies to the odd sites to deduce that

$$\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] \geq \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_O, A)\right]$$

□

The advantage of working with $\nu(\cdot \mid \sigma_O, A)$ or $\nu(\cdot \mid \sigma_E, A)$ instead of $\nu(\cdot \mid A)$ is that once we condition on the spins on all odd (resp. even) sites the measure becomes a product over the even (resp. odd) vertices, and we can exploit tensorization properties of entropy for product measures. The next lemma is a key step in the proof.

Lemma 4.3.4. *There exists a constant $\delta_1 > 0$ depending only on d, β, q such that, for all functions $f : \Omega_\mathrm{J} \mapsto \mathbb{R}_+$,*

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma)\right] + \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_O, A)\right] \geq \delta_1\, \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_O)\right], \tag{4.3.1}$$

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma)\right] + \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E, A)\right] \geq \delta_1\, \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E)\right]. \tag{4.3.2}$$

We defer the proof of Lemma 4.3.4 to later. Adding up (4.3.1) and (4.3.2) and using Lemma 4.3.3 we obtain the estimate

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma) + \mathrm{Ent}_\nu(f \mid A)\right] \geq \frac{\delta_1}{2}\, \nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E) + \mathrm{Ent}_\nu(f \mid \sigma_O)\right]. \tag{4.3.3}$$

We then use a generalization of the entropy factorization to reconstruct, in the presence of approximate even/odd factorization, the global entropy $\mathrm{Ent}_\nu(f)$ from the conditional average entropies $\nu\left[\mathrm{Ent}_\nu(f \mid \sigma_E)\right]$ and $\nu\left[\mathrm{Ent}_\nu(f \mid \sigma_O)\right]$ on the right hand side of (4.3.3).

Lemma 4.3.5. *Approximate even/odd factorization with consant C implies that for all*

functions $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$,

$$\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_E) + \mathrm{Ent}_\nu(f \,|\, \sigma_O)\right] \geq \delta_2 \mathrm{Ent}_\nu(f),$$

where $\delta_2 = 1/C$.

Proof. We need the following observations:

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O) = \mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right] \,|\, \sigma_O\right) + \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_O\right], \tag{4.3.4}$$

$$\mathrm{Ent}_\nu(f \,|\, \sigma_E) = \mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right] \,|\, \sigma_E\right) + \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_E\right]. \tag{4.3.5}$$

Indeed, to establish (4.3.4) note that from the definition of conditional entropy we get

$$\begin{aligned}
\mathrm{Ent}_\nu(f \,|\, \sigma_O) &= \nu\left[f \log\left(\frac{f}{\nu[f \,|\, \sigma_O]}\right) \,\middle|\, \sigma_O\right] \\
&= \nu\left[f \log\left(\frac{f}{\nu[f \,|\, \sigma_E, \sigma_O]}\right) \,\middle|\, \sigma_O\right] + \nu\left[f \log\left(\frac{\nu[f \,|\, \sigma_E, \sigma_O]}{\nu[f \,|\, \sigma_O]}\right) \,\middle|\, \sigma_O\right] \\
&= \nu\left[f \log\left(\frac{f}{\nu[f \,|\, \sigma_E, \sigma_O]}\right) \,\middle|\, \sigma_O\right] + \nu\left[\nu[f \mid \sigma] \log\left(\frac{\nu[f \,|\, \sigma]}{\nu[f \,|\, \sigma_O]}\right) \,\middle|\, \sigma_O\right] \\
&= \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_O\right] + \mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right] \,|\, \sigma_O\right),
\end{aligned}$$

where we also use the fact that $\nu[\cdot \,|\, \sigma_E, \sigma_O] = \nu[\cdot \,|\, \sigma]$. The same argument applies to (4.3.5).

Now, since the function $\nu\left[f \,|\, \sigma\right]$ depends only on the spin configuration σ,

$$\nu\left[\mathrm{Ent}_\nu(\nu[f \,|\, \sigma] \,|\, \sigma_E) + \mathrm{Ent}_\nu(\nu[f \,|\, \sigma] \,|\, \sigma_O)\right] = \mu\left[\mathrm{Ent}_\mu(\nu[f \,|\, \sigma] \,|\, \sigma_E) + \mathrm{Ent}_\mu(\nu[f \,|\, \sigma] \,|\, \sigma_O)\right];$$

and we may apply the approximate even/odd factorization to the function $\nu\left[f \,|\, \sigma\right]$. Then, there exists a constant $\delta_2 \in (0, 1]$ such that

$$\mu\left[\mathrm{Ent}_\mu(\nu[f \,|\, \sigma] \,|\, \sigma_E) + \mathrm{Ent}_\mu(\nu[f \,|\, \sigma] \,|\, \sigma_O)\right] \geq \delta_2 \,\mathrm{Ent}_\mu\left(\nu\left[f \,|\, \sigma\right]\right). \tag{4.3.6}$$

Therefore, observing that

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

we obtain from (4.3.4), (4.3.5) and (4.3.6)

$$\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_E) + \mathrm{Ent}_\nu(f \,|\, \sigma_O)\right] \geq \delta_2 \,\mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right]\right) + 2\,\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right].$$

Since $\delta_2 \leq 1$, the standard decomposition in (4.2.2) implies

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

as claimed. □

4.3.2 Proofs of main results

The proofs of Lemma 4.1.3 and Theorems 1.3.3 and 4.1.1 are now immediate.

Proof of Lemma 4.1.3. We note that inequality (4.3.3) and Lemma 4.3.5 are valid for any bipartite graph of bounded degree; the result follows by taking $C = 2/\delta_1\delta_2$. □

Proof of Theorem 4.1.1. It follows immediately from Lemmas 4.1.3 and 4.1.4. □

We now also prove Theorem 1.3.3.

Proof of Theorem 1.3.3. Theorem 3.3.2 implies that the even/odd factorization holds for any boundary condition whenever SSM holds. Then, from Lemma 4.3.2 we know that approximate spin/edge factorization holds; the result then follows from applying Lemmas 4.1.3 and 4.1.4. □

It remains for us to provide the proof of Lemma 4.3.4, which we do in the next subsection.

4.3.3 Proof of Lemma 4.3.4

Before giving the proof of Lemma 4.3.4, we mention several useful facts about the joint distribution ν. The first key fact is that, for any fixed configuration σ_O of spins on the odd sub-lattice, the conditional probability $\nu(\cdot \mid \sigma_O)$ is a product measure. That is,

$$\nu(\cdot \mid \sigma_O) = \bigotimes_{x \in E} \nu_x(\cdot \mid \sigma_O), \tag{4.3.7}$$

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

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

$$\nu(\cdot \,|\, \sigma_O, A) = \bigotimes_{x \in E} \nu_x(\cdot \,|\, \sigma_O, A),$$

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

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 $A_e = 0$ on all remaining edges. Therefore we may write

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

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

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

Lemma 4.3.6. *There exists a constant $\delta_1 > 0$ such that, for all functions $f \geq 0$ and all $x \in E$,*

$$\nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma) \,|\, \sigma_O \right] + \nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma_O, A) \,|\, \sigma_O \right] \geq \delta_1 \, \mathrm{Ent}_x(f \,|\, \sigma_O). \tag{4.3.8}$$

Proof. For $x \in V$, let A_x be random variable in $\{0,1\}^{\deg(x)}$ corresponding to the configuration of the edges incident to x in A. If we replace entropy by variance, then (4.3.8) is a spectral gap inequality for the Markov chain where the variable $(\sigma_x, A_x) \in [q] \times \{0,1\}^{\deg(x)} =: \mathcal{S}$ is updated as follows. At each step, with probability $1/2$ the spin σ_x is updated with a sample from $\nu_x(\cdot \,|\, \sigma_O, A)$, and with probability $1/2$ the edges A_x incident to x are simultaneously updated with a sample from $\nu_x(\cdot \,|\, \sigma)$. Let $P_x = \frac{Q_x + S_x}{2}$ denote the transition matrix of this Markov chain, where Q_x, S_x are the stochastic matrices corresponding to the spin and edge moves at x, respectively. Let $\mathcal{D}_{P_x}$, $\mathcal{D}_{Q_x}$ and $\mathcal{D}_{S_x}$ denote the corresponding Dirichlet forms. Observe that, by updating first the edges with an empty configuration and then the spin, two arbitrary initial configurations can be coupled after two steps with

probability at least $\frac{1}{4}(1-p)^{-2d}$, and thus for any function $f : \mathcal{S} \mapsto \mathbb{R}_+$

$$\frac{\mathcal{D}_{Q_x}(f,f) + \mathcal{D}_{S_x}(f,f)}{2} = \mathcal{D}_{P_x}(f,f) \geq \delta_0 \operatorname{Var}_x(f \,|\, \sigma_O),$$

where $\delta_0 > 0$ is a constant depending only on p and d. Using the standard facts that $\mathcal{D}_{Q_x}(f,f) = \nu_x \left[\operatorname{Var}_x(f \,|\, \sigma) \,|\, \sigma_O\right]$ and $\mathcal{D}_{S_x}(f,f) = \nu_x \left[\operatorname{Var}_x(f \,|\, \sigma_O, A) \,|\, \sigma_O\right]$, we arrive at the inequality

$$\frac{\nu_x \left[\operatorname{Var}_x(f \,|\, \sigma) \,|\, \sigma_O\right] + \nu_x \left[\operatorname{Var}_x(f \,|\, \sigma_O, A) \,|\, \sigma_O\right]}{2} \geq \delta_0 \operatorname{Var}_x(f \,|\, \sigma_O). \tag{4.3.9}$$

A well known general relation between entropy and variance (see, e.g., Theorem A.1 and Corollary A.4 in [47]) shows that, for all $f \geq 0$,

$$\operatorname{Ent}_x(f \,|\, \sigma_O) \leq C_1 \operatorname{Var}_x(\sqrt{f} \,|\, \sigma_O), \tag{4.3.10}$$

where $C_1 = C_1(q,p,d)$ is a constant independent of n, since we are considering the conditional measure at the single site x. Thus, applying (4.3.9) to $\sqrt{f}$ instead of f, we obtain

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

The conclusion (4.3.8) follows by recalling that for any $f \geq 0$ the variance of $\sqrt{f}$ is at most the entropy of f for any underlying probability measure; see, e.g., [84, Lemma 1]. In particular, $\operatorname{Var}_x(\sqrt{f} \,|\, \sigma) \leq \operatorname{Ent}_x(f \,|\, \sigma)$ and $\operatorname{Var}_x(\sqrt{f} \,|\, \sigma_O, A) \leq \operatorname{Ent}_x(f \,|\, \sigma_O, A)$. □

To prove Lemma 4.3.4, we need to lift the inequality of Lemma 4.3.6 to the product measure $\nu(\cdot \,|\, \sigma_O) = \otimes_{x \in E} \nu_x(\cdot \,|\, \sigma_O)$.

Proof of Lemma 4.3.4. We will prove (4.3.1); exactly the same argument applies to (4.3.2). Let $x = 1, \ldots, w$ denote an arbitrary ordering of the even sites $x \in E$. Let $A_x \in \{0,1\}^{\deg(x)}$ be the random variable corresponding to the state of the edges incident to x. We write $\xi_x = (\sigma_x, A_x)$ for the pair of variables at x. We first observe that

$$\operatorname{Ent}_\nu(f \,|\, \sigma_O) = \sum_{x=1}^{w} \nu \left[\operatorname{Ent}_x(g_{x-1} \,|\, \sigma_O) \,|\, \sigma_O\right], \tag{4.3.11}$$

where $g_x = \nu \left[f \,|\, \sigma_O, \xi_{x+1}, \ldots, \xi_w\right]$, so that $g_0 = f$ and $g_w = \nu \left[f \,|\, \sigma_O\right]$. To prove (4.3.11), we note that since $\nu(\cdot \,|\, \sigma_O) = \otimes_{x \in E} \nu_x(\cdot \,|\, \sigma_O)$, one has $\nu_x[g_{x-1} \,|\, \sigma_O] = g_x$.

Therefore,

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O) = \nu\left[g_0 \log\left(g_0/g_w\right) \,|\, \sigma_O\right] = \sum_{x=1}^{w} \nu\left[g_0 \log\left(g_{x-1}/g_x\right) \,|\, \sigma_O\right].$$

Since the g_x are (conditional) expectations, we deduce

$$\begin{aligned}
\mathrm{Ent}_\nu(f \,|\, \sigma_O) &= \sum_{x=1}^{w} \nu\left[g_{x-1} \log\left(g_{x-1}/g_x\right) \,|\, \sigma_O\right] \\
&= \sum_{x=1}^{w} \nu\left[\nu_x\left[g_{x-1} \log\left(g_{x-1}/g_x\right) \,|\, \sigma_O\right] \,|\, \sigma_O\right] \\
&= \sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O) \,|\, \sigma_O\right]. \qquad (4.3.12)
\end{aligned}$$

From (4.3.12), using Lemma 4.3.6 we obtain

$$\begin{aligned}
\delta_1 \mathrm{Ent}_\nu(f \,|\, \sigma_O) &\leq \sum_{x=1}^{w} \nu\left[\nu_x\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_O\right] + \nu_x\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right] \,|\, \sigma_O\right] \\
&= \sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) + \mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right]. \qquad (4.3.13)
\end{aligned}$$

Observe that $\sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_O\right]$ and $\sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right]$ are "tensorized" versions of $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_O\right]$ and $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_O, A)\right]$, respectively, which are the terms on the right hand side of (4.3.1). Using similar but somewhat more involved ideas to those used to derive (4.3.12), we can establish the following.

Lemma 4.3.7.

1. $\sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_O\right] \leq \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_O\right]$
2. $\sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right] \leq \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_O, A) \,|\, \sigma_O\right]$

Before providing the proof of this lemma, we finish the proof of Lemma 4.3.4. Inequality (4.3.13) together with parts 1 and 2 of Lemma 4.3.7 show that

$$\delta_1 \mathrm{Ent}_\nu(f \,|\, \sigma_O) \leq \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_O\right] + \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_O, A) \,|\, \sigma_O\right]. \qquad (4.3.14)$$

Taking expectations with respect to ν in (4.3.14) we arrive at (4.3.1) and the proof is complete. □

We finish the proof of Lemma 4.3.4 by providing the proof of Lemma 4.3.7.

Proof of Lemma 4.3.7. We start with part 2. Let $h_x = \nu\left[f \,|\, \sigma_O, \sigma_{x+1}, \ldots, \sigma_w, A\right]$, so

that $h_0 = f$ and $h_w = \nu\,[f \,|\, \sigma_O, A]$. Since $\nu(\cdot \,|\, \sigma_O, A)$ is a product measure,

$$\nu_x[h_{x-1} \,|\, \sigma_O, A] = h_x.$$

Therefore, reasoning as in (4.3.11) we obtain

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O, A) = \sum_{x=1}^{w} \nu\left[\mathrm{Ent}_x(h_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O, A\right]. \tag{4.3.15}$$

Taking expectations with respect to $\nu(\cdot \,|\, \sigma_O)$ in (4.3.15) we see that it is sufficient to show that, for all x,

$$\nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right] \leq \nu\left[\mathrm{Ent}_x(h_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right]. \tag{4.3.16}$$

To prove (4.3.16), we introduce the measures

$$\mu_k = \otimes_{x=1}^{k} \nu_x(\cdot \,|\, \sigma_O)$$

and

$$\mu_k^A = \otimes_{x=1}^{k} \nu_x(\cdot \,|\, \sigma_O, A).$$

Then we have $g_x = \mu_x[f]$, $h_x = \mu_x^A[f]$, and $g_x = \mu_x[h_x]$. Also, we simplify the notation by writing $\nu_x(\cdot \,|\, \sigma_O, A) =: \nu_x^A$. Now the product structure implies the commutation relation between expectations

$$\nu_x^A g_{x-1} = \nu_x^A \mu_{x-1} h_{x-1} = \mu_{x-1} \nu_x^A h_{x-1}. \tag{4.3.17}$$

Therefore,

$$\begin{aligned}
\nu\left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_O, A) \,|\, \sigma_O\right] &= \nu\left[g_{x-1} \log\left(g_{x-1}/\nu_x^A g_{x-1}\right) \,|\, \sigma_O\right] \\
&= \nu\left[\mu_{x-1}h_{x-1} \log\left(\mu_{x-1}h_{x-1}/\mu_{x-1}\nu_x^A h_{x-1}\right) \,|\, \sigma_O\right] \\
&= \nu\left[h_{x-1} \log\left(\mu_{x-1}h_{x-1}/\mu_{x-1}\nu_x^A h_{x-1}\right) \,|\, \sigma_O\right] \\
&= \nu\left[\nu_x^A\left(h_{x-1} \log\left(g_{x-1}/\nu_x^A g_{x-1}\right)\right) \,|\, \sigma_O\right].
\end{aligned} \tag{4.3.18}$$

From the variational principle (2.3.2) it follows that

$$\nu_x^A\left[h_{x-1} \log\left(g_{x-1}/\nu_x^A[g_{x-1}]\right)\right] \leq \mathrm{Ent}_x(h_{x-1} \,|\, \sigma_O, A), \tag{4.3.19}$$

which combined with (4.3.18) proves (4.3.16). This completes the proof of part 2. We use a similar argument for part 1. Let $\psi_x = \nu\,(f \,|\, \sigma, A_{x+1}, \ldots, A_w)$, so that

$\psi_0 = f$ and $\psi_w = \nu(f \mid \sigma)$. Notice that $\nu_x[\psi_{x-1} \mid \sigma] = \psi_x$. Therefore, as in (4.3.11),

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

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

$$\nu\left[\mathrm{Ent}_x(g_{x-1} \mid \sigma) \mid \sigma_O\right] \leq \nu\left[\mathrm{Ent}_x(\psi_{x-1} \mid \sigma) \mid \sigma_O\right]. \tag{4.3.20}$$

Introducing the measures $\mu_k = \otimes_{x=1}^k \nu_x(\cdot \mid \sigma_O)$, $\mu_k^\sigma = \otimes_{x=1}^k \nu_x(\cdot \mid \sigma)$, and $\nu_x^\sigma = \nu_x(\cdot \mid \sigma)$, we have $g_x = \mu_x[f]$, $\psi_x = \mu_x^\sigma[f]$, and $g_x = \mu_x[\psi_x]$. As in (4.3.17), we have the commutation relation

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

Therefore, as in (4.3.18)-(4.3.19) we obtain

$$\begin{aligned}
\nu\left[\mathrm{Ent}_x(g_{x-1} \mid \sigma) \mid \sigma_O\right] &= \nu\left[g_{x-1}\log\left(g_{x-1}/\nu_x^\sigma g_{x-1}\right) \mid \sigma_O\right] \\
&= \nu\left[\mu_{x-1}\psi_{x-1}\log\left(\mu_{x-1}\psi_{x-1}/\mu_{x-1}\nu_x^\sigma\psi_{x-1}\right) \mid \sigma_O\right] \\
&= \nu\left[\psi_{x-1}\log\left(\mu_{x-1}\psi_{x-1}/\mu_{x-1}\nu_x^\sigma\psi_{x-1}\right) \mid \sigma_O\right] \\
&= \nu\left[\nu_x^\sigma\left(\psi_{x-1}\log\left(g_{x-1}/\nu_x^\sigma g_{x-1}\right)\right) \mid \sigma_O\right] \\
&\leq \nu\left[\mathrm{Ent}_x(\psi_{x-1} \mid \sigma) \mid \sigma_O\right].
\end{aligned}$$

This proves (4.3.20) and completes the proof of part 1. □

4.4 A lower bound for the SW dynamics

In this section we establish an asymptotically tight lower bound for the mixing time of SW dynamics whenever SSM holds; this result implies the lower bound in Theorem 1.3.3 from the introduction.

Theorem 4.4.1. *In an n-vertex cube of $\mathbb{Z}^d$, for all integer $q \geq 2$ and all $\beta > 0$, SSM implies that for all boundary conditions $T_{\mathrm{mix}}(SW) = \Omega(\log n)$.*

The main new ingredient in the proof of this result is a bound on the speed of propagation of disagreements under a coupling of the steps of the SW dynamics provided SSM holds. With this new tool, we are able to adapt the lower bound framework of Hayes and Sinclair [77] for the Glauber dynamics to the SW setting. We also use a recently established fact about concentration properties of the Potts measure due to [51].

SW coupling. Consider two copies of the SW dynamics on the graph $G = (V, \mathbb{E})$, where V is an n-vertex cube of $\mathbb{Z}^d$. Let X_t and Y_t be the configurations of these copies at time $t \geq 0$. We can couple the steps of the SW dynamics as follows:

1. Draw $|\mathbb{E}|$ independent, uniform random numbers from $[0, 1]$, one for each edge. Let $r_e(t) \in [0, 1]$ denote the random number corresponding to the edge $e \in \mathbb{E}$.

2. Draw $|V|$ independent, uniform random numbers from $\{1, ..., q\}$, one for each vertex. Let $s_v(t) \in \{1, ..., q\}$ denote the random number for $v \in V$.

3. Let $A_X = \{e \in M(X_t) : r_e(t) \leq p\}$ and $A_Y = \{e \in M(Y_t) : r_e(t) \leq p\}$, where recall that $M(X_t)$ and $M(Y_t)$ denote the set of monochromatic edges in X_t and Y_t, respectively

4. For each connected component $\mathcal{C}$ of (V, A_X) or (V, A_Y), we let $s_{\mathcal{C}} = s_v(t)$, where v is the vertex in $\mathcal{C}$ with the smallest coordinate sum. (If two or more vertices in $\mathcal{C}$ have the same coordinate sum, we break ties "lexicographically" using the coordinates.) Then, every vertex of $\mathcal{C}$ is assigned the spin $s_{\mathcal{C}}$.

The key property of the SW coupling is that, after assigning the edges, two identical connected components in A_X and A_Y will be assigned the same spin (namely, the spin s_v of their common vertex v with smallest coordinate sum). We show that, under SSM, the SW coupling propagates disagreements slowly for a suitable starting condition.

To describe our starting condition we introduce the notion of *L-shattered* configurations.

Definition 4.4.2. Consider the graph $G = (V, \mathbb{E})$, where V is an n-vertex cube of $\mathbb{Z}^d$. For a configuration σ on V, let $A_\sigma \subseteq M(\sigma)$ be the configuration that results from keeping each monochromatic edge in $M(\sigma)$ independently with probability $p = 1 - \exp(-\beta)$. We say that σ is *L-shattered* in V if, with probability at least $1 - |V| \exp(-\gamma L)$ where $\gamma > 0$ is a fixed constant we choose later, for every $v \in |V|$ at distance at least $2L$ from the boundary of $|V|$, the connected component of v in A_σ does not reach the boundary of the cube $\Lambda_v(L)$ centered at v of side length L.

Note that the above defined notion involves a probability that decays exponentially with L, so the dimension of the cube V will not be as significant as long as $\log n \ll L \ll n$. The following lemma establishes a concentration of the probability mass on shattered configurations under SSM (for the monochromatic "all 1" boundary condition).

Lemma 4.4.3. *Let $\mathfrak{S}$ be the set of L-shattered configurations of the n-vertex cube $G = (V, \mathbb{E})$ of $\mathbb{Z}^d$. There exists a constant $c > 0$ such that for all integers $q \geq 2$ and $L \geq 1$, SSM implies that $\mu^1(\mathfrak{S}) \geq 1 - \exp(-cL)$.*

The proof of this lemma, which follows straightforwardly from the results in [51], will be provided later in Section 4.4.1. We can now describe our starting condition for the SW dynamics.

A starting condition. We consider a regular pattern of non-overlapping cubes in $\mathbb{Z}^d$ of side length $\ell = (\log n)^3$ with a fixed minimal distance between cubes. Formally, consider the cubes of side length ℓ centered at $(\ell + 3) \cdot \vec{h}$ where $\vec{h} \in \mathbb{Z}^d$. These cubes have volume ℓ^d and are at distance 4 from each other. We let $B_1, B_2, \ldots B_N \subset V$ be the collection of those cubes that are contained in V and at distance at least 4 from the boundary ∂V; then, $N = \Theta(n/\ell^d)$.

Let $B = \bigcup_{i=1}^N B_i$, $\partial B = \bigcup_{i=1}^N \partial B_i$ and let e_i be an edge at the center of B_i. For definiteness, we may assume that ℓ is odd so that there is a unique vertex v_i at the center of each B_i; we take $e_i = \{v_i, u_i\}$ where $u_i = v_i + (1, 0, \ldots, 0) \in \mathbb{Z}^d$. Let $\mathcal{A}_i$ be the set of configurations on B_i in which the spins at the endpoints of e_i are the same, and let $\mathcal{S}_i$ be the set of L-shattered configurations in each B_i. (Later we will set $L = C(\log n)$ with $C > 0$ a large constant.)

We consider two variants of the SW dynamics, $\{X_t\}$ and $\{Y_t\}$, with the same initial condition $X_0 = Y_0$. The chain $\{X_t\}$ is an instance of the standard SW dynamics on $(V, \mathbb{E})$; for the initial state X_0 of $\{X_t\}$ we set the spins of all the vertices in $U = (V \setminus B) \cup \partial B$ to 1. The configuration in each cube B_i is sampled (independently) proportional to $\mu^1_{B_i}$ on $\mathcal{S}_i \cap \mathcal{A}_i$, where $\mu^1_{B_i}$ denotes the Potts measure on B_i with the "all 1" monochromatic boundary condition.

The other instance we consider, $\{Y_t\}$, only updates the spins of the vertices in $B \setminus \partial B$. That is, after adding all the monochromatic edges independently with probability $p = 1 - \exp(-\beta)$, only the connected components fully contained in B update their spins. (Note that if a component touches the boundary of B, then it is not updated since the boundary is frozen to the spin 1 by the boundary condition.) We set $Y_0 = X_0$ and couple the evolution of Y_t and X_t using the SW coupling defined earlier. We can view $\{Y_t\}$ as a dynamics on the configurations on B whose stationary measure is $\mu^1_B = \otimes_{i=1}^N \mu^1_{B_i}$. We also observe that a step of $\{Y_t\}$ is equivalent to performing one step of the SW dynamics in each B_i independently.

Note that $X_0 = Y_0$, and any disagreements between X_t, Y_t at later times t can arise only from the fact that Y_t does not update the spins outside B: i.e., disagreements must propagate into the B_i from their boundaries. The following result, whose proof we defer until after the proof of Theorem 4.4.1, provides a bound on the speed of propagation of these disagreements under the SW coupling with the

specified initial condition. In particular it says that, for $t = \Omega(\log n)$ steps, X_t, Y_t agree w.h.p. on the spins at the center of every cube B_i.

Theorem 4.4.4. *Let $\mathcal{C} = \bigcup_{i=1}^{N} e_i$ and set $L = C(\log n)$. For any constant $A > 0$, for a sufficiently large constant $C > 0$ SSM implies that*

$$\Pr\left[\forall t \leq A \log n : \ X_t(\mathcal{C}) = Y_t(\mathcal{C})\right] = 1 - o(1).$$

A key ingredient in the proof of Theorem 4.4.4 (and also of Theorem 4.4.1) is the following discrete time version of the completely monotone decreasing (CMD) property of reversible Markov chains from [77]; the proof of this lemma is provided in Section 4.4.1.

Lemma 4.4.5. *Let $\{X_t\}$ denote a discrete time Markov chain with finite state space Ω, reversible with respect to π and with a positive semidefinite transition matrix. Let $B \subset \Omega$ denote an event. If X_0 is sampled proportional to π on B, then $\Pr(X_t \in B) \geq \pi(B)$ for all $t \geq 0$, and for all $t \geq 1$*

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

We now proceed with the proof of Theorem 4.4.1.

Proof of Theorem 4.4.1. Our goal is to show that at some time $T = \Theta(\log n)$

$$\|X_T - \mu\|_{\text{TV}} > \frac{1}{2},$$

where with a slight abuse of notation we use X_T for the distribution of the chain $\{X_t\}$ at time T. This clearly implies that the mixing time of the SW dynamics is $\Omega(\log n)$.

Let $\mathcal{C} = \bigcup e_i$ and let $e_i = \{a_i, b_i\}$. Let $\hat{\mu}_{\mathcal{C}}$ and $\hat{\mu}^1_{\mathcal{C}}$ be the marginals of μ and μ^1_B, respectively, on $\mathcal{C}$. Then,

$$\begin{aligned}
\|X_T - \mu\|_{\text{TV}} &\geq \|X_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}\|_{\text{TV}} \\
&\geq \|Y_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}\|_{\text{TV}} - \|X_T(\mathcal{C}) - Y_T(\mathcal{C})\|_{\text{TV}} \\
&\geq \|Y_T(\mathcal{C}) - \hat{\mu}^1_{\mathcal{C}}\|_{\text{TV}} - \|\hat{\mu}^1_{\mathcal{C}} - \hat{\mu}_{\mathcal{C}}\|_{\text{TV}} - \|X_T(\mathcal{C}) - Y_T(\mathcal{C})\|_{\text{TV}}. \quad (4.4.1)
\end{aligned}$$

We bound each term of (4.4.1) independently. We note first that by Theorem 4.4.4

$$\|X_T(\mathcal{C}) - Y_T(\mathcal{C})\|_{\text{TV}} \leq \Pr(X_T(\mathcal{C}) \neq Y_T(\mathcal{C})) = o(1).$$

We proceed to bound the term $\|\hat{\mu}^1_{\mathcal{C}} - \hat{\mu}_{\mathcal{C}}\|_{\text{TV}}$ in (4.4.1), for which we use SSM. Let $\Omega(A)$ be the set of all possible configurations on the set $A \subseteq V$. For a configu-

ration ψ on U, let $\hat{\mu}_{\mathcal{C}}^{\psi}$ denote the marginal of μ_B^{ψ} on $\mathcal{C}$. Let $\hat{\mu}_{e_i}^1$, $\hat{\mu}_{e_i}^{\psi}$ be the marginals of $\hat{\mu}_{B_i}^1$, $\hat{\mu}_{B_i}^{\psi}$ on e_i, respectively. Then,

$$\|\hat{\mu}_{\mathcal{C}}^1 - \hat{\mu}_{\mathcal{C}}\|_{\text{TV}} \leq \sum_{\psi\in\Omega(U)} \mu(\psi)\|\hat{\mu}_{\mathcal{C}}^1 - \hat{\mu}_{\mathcal{C}}^{\psi}\|_{\text{TV}} \leq \sum_{\psi\in\Omega(U)} \sum_{i=1}^{N} \mu(\psi)\|\hat{\mu}_{e_i}^1 - \hat{\mu}_{e_i}^{\psi}\|_{\text{TV}} \leq \frac{N}{e^{\kappa\ell}} = o(1),$$

where the second inequality follows from the fact that μ_B^1 and μ_B^{ψ} are product measures over the B_i's, and the last one follows from the SSM property for a suitable constant $\kappa > 0$.

It remains for us to provide a lower bound for the term $\|Y_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}^1\|_{\text{TV}}$ in (4.4.1). For this, we introduce an auxiliary copy of the chain $\{Y_t\}$, denoted $\{Z_t\}$, which is coupled with $\{Y_t\}$ but with a slightly different starting condition. Namely, Z_0 is sampled proportional to $\mu_{B_i}^1$ on the set $\mathcal{A}_i$, independently for each B_i. (Recall that $Y_0 = X_0$ is sampled proportional to $\mu_{B_i}^1$ on $\mathcal{S}_i \cap \mathcal{A}_i$ instead.) Then,

$$\|Y_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}^1\|_{\text{TV}} \geq \|Z_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}^1\|_{\text{TV}} - \|Y_T(\mathcal{C}) - Z_T(\mathcal{C})\|_{\text{TV}}. \tag{4.4.2}$$

We first provide an upper bound for the second term in (4.4.2). Plainly,

$$\|Y_T(\mathcal{C}) - Z_T(\mathcal{C})\|_{\text{TV}} \leq \|Y_T - Z_T\|_{\text{TV}} \leq \Pr[Y_0 \neq Z_0].$$

Let μ_0^Y, μ_0^Z be the initial distribution for $\{Y_t\}$ and $\{Z_t\}$, respectively, and let $\mathcal{S} = \otimes\mathcal{S}_i$ and $\mathcal{A} = \otimes\mathcal{A}_i$. For $\sigma \in \mathcal{S} \cap \mathcal{A}$, we have $\mu_0^Y(\sigma) = \mu_B^1(\sigma)/\mu_B^1(\mathcal{S} \cap \mathcal{A})$, and for $\sigma \in \mathcal{A}$, $\mu_0^Z(\sigma) = \mu_B^1(\sigma)/\mu_B^1(\mathcal{A})$. Therefore, if the configurations Y_0 and Z_0 are sampled from the optimal coupling between μ_0^Y, μ_0^Z and the steps of $\{Y_t\}$, $\{Z_t\}$ are then coupled with the SW coupling, we have

$$\|Y_T(\mathcal{C}) - Z_T(\mathcal{C})\|_{\text{TV}} \leq \|\mu_0^Y - \mu_0^Z\|_{\text{TV}} \leq \sum_{i=1}^{N} \|\mu_0^{Y,i} - \mu_0^{Z,i}\|_{\text{TV}} = N\|\mu_0^{Y,1} - \mu_0^{Z,1}\|_{\text{TV}},$$

where $\mu_0^{Y,i}$, $\mu_0^{Z,i}$ are the initial distributions of Y_0, Z_0 on B_i. Then,

$$\|\mu_0^{Y,1} - \mu_0^{Z,1}\|_{\text{TV}} = \frac{\mu_{B_1}^1(\mathcal{A}_1 \setminus \mathcal{S}_1)}{\mu_{B_1}^1(\mathcal{A}_1)} \leq \frac{\mu_{B_1}^1(\mathcal{S}_1^c)}{\mu_{B_1}^1(\mathcal{A}_1)} = O\left(e^{-cL}\right),$$

where the last inequality follows from Lemma 4.4.3 and the fact that $\mu_{B_1}^1(\mathcal{A}_1) = \Omega(1)$. In summary, since $L = C(\log n)$ and C can be taken large enough, we have proved

$$\|Y_T(\mathcal{C}) - Z_T(\mathcal{C})\|_{\text{TV}} = o(1).$$

It remains for us to find a lower bound for $\|Z_T(\mathcal{C}) - \hat{\mu}_{\mathcal{C}}^1\|_{\text{TV}}$ in (4.4.2) for a suitable T. For a configuration σ on B, let $f(\sigma)$ denote the number of edges $e_i \in \mathcal{C}$

that are monochromatic in σ. For any $a \geq 0$ we have

$$\|Z_T(\mathcal{C}) - \hat{\mu}^1_{\mathcal{C}}\|_{\text{TV}} \geq \Pr[f(Z_T) \geq a] - \Pr_{\sigma\sim\mu^1_B}[f(\sigma) \geq a]. \tag{4.4.3}$$

We will show that, for a suitable T and any $i = 1, \ldots, N$,

$$\Pr[Z_T(B_i) \in \mathcal{A}_i] \geq \mu^1_{B_i}(\mathcal{A}_i) + \frac{1}{N^{1/4}}. \tag{4.4.4}$$

Assuming this is the case, then setting $\mathcal{W} = \sum_{i=1}^N \mu^1_{B_i}(\mathcal{A}_i)$ we obtain by Hoeffding's inequality

$$\Pr\left[f(Z_T) \geq \mathcal{W} + N^{3/4} - \sqrt{N \log N}\right] \geq 1 - \frac{1}{N^2}$$

and

$$\Pr_{\sigma\sim\mu^1_B}\left[f(\sigma) \geq \mathcal{W} + \sqrt{N \log N}\right] \leq \frac{1}{N^2},$$

which yields from (4.4.3) that $\|Z_T(\mathcal{C}) - \hat{\mu}^1_{\mathcal{C}}\|_{\text{TV}} \geq 1 - 2/N^2$ by taking, e.g., $a = \mathcal{W} + \sqrt{N \log N}$.

To establish (4.4.4), note that by Lemma 4.4.5

$$\Pr(Z_T(B_i) \in \mathcal{A}_i) \geq \mu^1_{B_i}(\mathcal{A}_i) + (1 - \mu^1_{B_i}(\mathcal{A}_i))^{-T+1}(\Pr(Z_1(B_i) \in \mathcal{A}_i) - \mu^1_{B_i}(\mathcal{A}_i))^T. \tag{4.4.5}$$

We remark that $\{Z_t\}$ has positive semidefinite transition matrix; this follows from the fact $\{Z_t\}$ is a product of SW dynamics in each B_i, and the SW dynamics has positive semidefinite transition matrix [13].

Let $P^{(i)}_{\text{SW}}$ denote the transition matrix of the SW dynamics on B_i. Then

$$\begin{aligned}\Pr(Z_1(B_i) \in \mathcal{A}_i) &= \sum_{\sigma\in\mathcal{A}_i} \frac{\mu^1_{B_i}(\sigma)}{\mu^1_{B_i}(\mathcal{A}_i)} P^{(i)}_{\text{SW}}(\sigma, \mathcal{A}_i) = \sum_{\sigma\in\mathcal{A}_i} \frac{\mu^1_{B_i}(\sigma)}{\mu^1_{B_i}(\mathcal{A}_i)} \left(\theta(\sigma) + \frac{1-\theta(\sigma)}{q}\right) \\ &= \frac{1}{q} + \frac{q-1}{q\mu^1_{B_i}(\mathcal{A}_i)} \sum_{\sigma\in\mathcal{A}_i} \mu^1_{B_i}(\sigma)\theta(\sigma)\ , \end{aligned} \tag{4.4.6}$$

where $\theta(\sigma)$ denotes the probability that, after the edge percolation phase of the SW step, the end points of the edge e_i are connected in the edge configuration.

Similarly,

$$\begin{aligned}\mu^1_{B_i}(\mathcal{A}_i) &= \sum_{\sigma\in\Omega(B_i)} \mu^1_{B_i}(\sigma)P^{(i)}_{\mathrm{SW}}(\sigma,\mathcal{A}_i)\\ &= \sum_{\sigma\in\Omega(B_i)\setminus\mathcal{A}_i} \mu^1_{B_i}(\sigma)P^{(i)}_{\mathrm{SW}}(\sigma,\mathcal{A}_i) + \sum_{\sigma\in\mathcal{A}_i} \mu^1_{B_i}(\sigma)P^{(i)}_{\mathrm{SW}}(\sigma,\mathcal{A}_i)\\ &= \sum_{\sigma\in\Omega(B_i)\setminus\mathcal{A}_i} \frac{\mu^1_{B_i}(\sigma)}{q} + \sum_{\sigma\in\mathcal{A}_i} \mu^1_{B_i}(\sigma)\left(\theta(\sigma)+\frac{1-\theta(\sigma)}{q}\right)\\ &= \frac{1}{q} + \frac{q-1}{q}\sum_{\sigma\in\mathcal{A}_i}\mu^1_{B_i}(\sigma)\theta(\sigma).\end{aligned}$$

Combining with (4.4.6) we get

$$\begin{aligned}\Pr(Z_1(B_i)\in\mathcal{A}_i) - \mu^1_{B_i}(\mathcal{A}_i) &= \frac{q-1}{q}\left(\frac{1}{\mu^1_{B_i}(\mathcal{A}_i)}-1\right)\sum_{\sigma\in\mathcal{A}_i}\mu^1_{B_i}(\sigma)\theta(\sigma)\\ &\geq \frac{q-1}{q}\left(\frac{1}{\mu^1_{B_i}(\mathcal{A}_i)}-1\right)p\cdot\mu^1_{B_i}(\mathcal{A}_i)\\ &= \frac{q-1}{q}\left(1-\mu^1_{B_i}(\mathcal{A}_i)\right)p,\end{aligned}$$

where in the last inequality we use the fact that $\theta(\sigma)\geq p$ when $\sigma\in\mathcal{A}_i$; recall that $p=1-e^{-\beta}$.

Plugging this bound into (4.4.5), we obtain

$$\begin{aligned}\Pr(Z_T(B_i)\in\mathcal{A}_i) &\geq \mu^1_{B_i}(\mathcal{A}_i) + (1-\mu^1_{B_i}(\mathcal{A}_i))^{-T+1}\left(\frac{q-1}{q}\left(1-\mu^1_{B_i}(\mathcal{A}_i)\right)p\right)^T\\ &= \mu^1_{B_i}(\mathcal{A}_i) + (1-\mu^1_{B_i}(\mathcal{A}_i))\left(\frac{(q-1)p}{q}\right)^T \geq \mu^1_{B_i}(\mathcal{A}_i)+\frac{1}{N^{1/4}},\end{aligned}$$

where the last inequality holds for $T=\xi\log n$ for a suitable constant $\xi>0$ since $\mu^1_{B_i}(\mathcal{A}_i)=\Omega(1)$. □

We provide next the proof of Theorem 4.4.4, our bound on the speed of disagreement propagation under the SW coupling.

Proof of Theorem 4.4.4. We will show inductively that with high probability disagreements propagate a distance of at most L in each step. Let $\Lambda_i(k)\subseteq B_i$ be the cube of side length $k<\ell$ centered at v_i; recall that $e_i=\{v_i,u_i\}$ where v_i is the center of B_i. Let $\Lambda(k)=\bigcup_{i=1}^N\Lambda_i(k)$. Note that at time 0, X_0 and Y_0 agree on $B=\Lambda(\ell)$.

Let us assume that X_t and Y_t agree on $\Lambda(k)$ for some $k\leq\ell-2L$. Suppose Y_t is L-shattered in each B_i; i.e., $Y_t(B_i)\in\mathcal{S}_i$ for $i=1,\dots,N$. If $E(k)$ is the set of edges with both endpoints in $\Lambda(k)$, after adding the monochromatic edges of $E(k)$

in X_t and Y_t coupled with the SW coupling, the joint edge/spin configuration on $(\Lambda(k), E(k))$ will be the same in both copies. However, when assigning the new spins, the connected components are not necessarily the same since there can be external connections; i.e., monochromatic paths in $V \setminus \Lambda(k)$. This may create disagreements between the two chains on $\Lambda(k)$ but only in the components touching the boundary of $\Lambda(k)$. Since we are assuming that Y_t is L-shattered in each B_i, then with probability $1 - N|B_i| \cdot \exp(-\gamma L)$, the disagreements cannot propagate to $\Lambda(k - 2(L+1))$. Consequently, the spin configurations of X_{t+1} and Y_{t+1} on $\Lambda(k - 2(L+1))$ are the same.

Proceeding inductively, and assuming that Y_t is L-shattered in each B_i for all $t = 0, \ldots, T$, we deduce from a union bound that X_T and Y_T agree on $\Lambda(\ell - 2L - 2T(L+1))$ with probability at least $1 - TN|B_i|\exp(-\gamma L)$, provided $\ell > 2T(L+1)+2L$. Therefore, $X_t(\mathcal{C}) = Y_t(\mathcal{C})$ for all $t \le T$ since $\mathcal{C} \subseteq \Lambda(\ell - 2L - 2t(L+1))$.

It remains for us to show that Y_t is L-shattered in each B_i for all $t = 0, \ldots, T$ with probability at least $1 - o(1)$. The configuration of Y_0 on B_i is sampled proportional to $\mu^1_{B_i}$ on $\mathcal{S}_i \cap \mathcal{A}_i$. For $\sigma \in \mathcal{S}_i \cap \mathcal{A}_i$, let $\pi_i(\sigma) = \mu^1_{B_i}(\sigma)/\mu^1_{B_i}(\mathcal{S}_i \cap \mathcal{A}_i)$ and for $\sigma \in \mathcal{S}_i$, let $\hat{\pi}_i(\sigma) = \mu^1_{B_i}(\sigma)/\mu^1_{B_i}(\mathcal{S}_i)$. We have

$$\begin{aligned}
&\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i) \\
&\quad = \Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i \mid Y_0(B_i) \in \mathcal{A}_i)\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_0(B_i) \in \mathcal{A}_i) + \\
&\qquad \Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i \mid Y_0(B_i) \notin \mathcal{A}_i)\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_0(B_i) \notin \mathcal{A}_i),
\end{aligned}$$

and so

$$\begin{aligned}
\Pr_{Y_0(B_i)\sim\pi_i}(Y_t(B_i) \in \mathcal{S}_i) &= \Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i \mid Y_0(B_i) \in \mathcal{A}_i) \\
&\ge 1 - \frac{1 - \Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i)}{\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_0(B_i) \in \mathcal{A}_i)} \qquad (4.4.7)
\end{aligned}$$

By Lemmas 4.4.5 and 4.4.3,

$$\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_t(B_i) \in \mathcal{S}_i) \ge \mu^1_{B_i}(\mathcal{S}_i) \ge 1 - \frac{1}{e^{cL}}.$$

Moreover, $\Pr_{Y_0(B_i)\sim\hat{\pi}_i}(Y_0(B_i) \in \mathcal{A}_i) = \alpha(\beta, q, d) = \Omega(1)$, and so we obtain from (4.4.7)

$$\Pr_{Y_0(B_i)\sim\pi_i}(Y_t(B_i) \in \mathcal{S}_i) = 1 - O\left(\frac{1}{e^{cL}}\right).$$

Setting $\mathcal{S} = \otimes_{i=1}^N \mathcal{S}_i$, a union bound over the B_i's implies

$$\Pr(Y_t \in \mathcal{S}) = 1 - O\left(\frac{N}{e^{cL}}\right).$$

It follows from another union bound over the steps that $\Pr(\forall t \leq T : Y_t \in \mathcal{S}) \geq 1 - O\left(\frac{TN}{e^{cL}}\right)$. Setting $T = A(\log n)$ (which satisfies $\ell > 2T(L+1) + 2L$ as required), recalling that $N = \Theta(n/\ell^d)$, and taking C sufficiently large, we obtain that $\Pr(\forall t \leq T : Y_t \in \mathcal{S}) \geq 1 - o(1)$, and hence

$$\Pr\left[\forall t \leq A \log n : \ X_t(\mathcal{C}) = Y_t(\mathcal{C})\right] = 1 - o(1),$$

as claimed. □

4.4.1 Proof of auxiliary lemmas

We conclude the section with the proofs of Lemma 4.4.5 and 4.4.3.

Proof of Lemma 4.4.5. From the spectral decomposition (see, e.g., [87]), one has

$$\Pr(X_t \in B) = \pi(B) + (1 - \pi(B)) \sum_{\ell=2}^{|\Omega|} \kappa_\ell \lambda_\ell^t, \tag{4.4.8}$$

where $\kappa_\ell \geq 0$ and $\sum_{\ell=2}^{|\Omega|} \kappa_\ell = 1$, and $\lambda_2, \ldots, \lambda_{|\Omega|}$ denote the non-negative eigenvalues of the transition matrix of the Markov chain except for the principal eigenvalue $\lambda_1 = 1$. In particular,

$$\Pr(X_1 \in B) - \pi(B) = (1 - \pi(B)) \sum_{\ell=2}^{|\Omega|} \kappa_\ell \lambda_\ell. \tag{4.4.9}$$

The convexity of the function $f(x) = x^t$ for $t \geq 1$, $x \geq 0$ and Jensen's inequality imply that

$$\Pr(X_t \in B) \geq \pi(B) + (1 - \pi(B)) \left(\sum_{\ell=2}^{|\Omega|} \kappa_\ell \lambda_\ell \right)^t.$$

From (4.4.9) it follows that $\sum_{\ell=2}^{|\Omega|} \kappa_\ell \lambda_\ell = (1 - \pi(B))^{-1}(\Pr(X_1 \in B) - \pi(B))$, and so for $t \geq 1$

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

Finally, observe also that from (4.4.8) it follows that $\Pr(X_t \in B) - \pi(B) \geq 0$. □

Proof of Lemma 4.4.3. Let $G = (V, \mathbb{E})$ be an n-vertex cube in $\mathbb{Z}^d$. Let $\mathcal{E}_L$ be the set of all edge configurations $A \subseteq \mathbb{E}$ such that for all $v \in V$ at distance at least $2L$ from partial ∂V, the connected component of v in A does not reach the boundary of the cube $\Lambda_v(L)$ centered at v of side length L. Consider the admissible boundary

condition $(1,1)$ of the joint space that is wired and all spins are 1; i.e., in Definition 4.3.1 we set $V_0 = \partial V$, $E_0 = \partial E$, $\psi = 1$ and $\varphi = 1$. From Theorem 1.2 in [51], we get that for a suitable constant $a > 0$

$$\nu^{(1,1)}(\mathcal{E}_L) \geq 1 - \frac{n}{e^{aL}}. \tag{4.4.10}$$

We remark that Theorem 1.2 from [51] is stated for the random-cluster measure with the wired boundary condition, but our statement for the joint measure ν is equivalent; see Section 4.7 below for a definition of the random-cluster measure and its boundary conditions. We also note that Theorem 1.2 from [51] requires a weaker (easier to satisfy) condition than SSM. Recall that $\mathfrak{S}$ is the set of L-shattered configurations of V. Then,

$$\begin{aligned}\nu^{(1,1)}(\mathcal{E}_L) &= \sum_{\sigma\in\mathfrak{S}} \mu^1(\sigma)\nu^{(1,1)}(\mathcal{E}_L \mid \sigma) + \sum_{\sigma\in\mathfrak{S}^c} \mu^1(\sigma)\nu^{(1,1)}(\mathcal{E}_L \mid \sigma) \\ &\leq \mu^1(\mathfrak{S}) + (1-\mu^1(\mathfrak{S}))\left(1 - \frac{n}{e^{\gamma L}}\right) = 1 - \frac{n}{e^{\gamma L}} + \frac{n\mu^1(\mathfrak{S})}{e^{\gamma L}}.\end{aligned}$$

Combining this with (4.4.10), we obtain that $\mu^1(\mathfrak{S}) \geq 1 - \exp(-(a-\gamma)L)$; the result follows by choosing $\gamma = a/2 = c$. □

4.5 Entropy decay for dynamics in the joint space

In this section we study the implications of our spin/edge factorization of entropy with respect to the joint measure ν for various dynamics on the joint space on $\mathbb{Z}^d$.

4.5.1 Swendsen-Wang in the joint space

First, we consider the SW dynamics in the joint space. Let K denote the $|\Omega_{\text{J}}| \times |\Omega_{\text{J}}|$ stochastic matrix corresponding to re-sampling the spins of a joint configuration given the edges, and similarly let T be the stochastic matrix corresponding to re-sampling the edges given the spins. Specifically,

$$\begin{aligned}K((\sigma,A),(\tau,B)) &= \mathbf{1}(A=B)\nu(\tau \mid A) \\ T((\sigma,A),(\tau,B)) &= \mathbf{1}(\sigma=\tau)\nu(B \mid \sigma).\end{aligned}$$

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

The Markov chains with transition matrices KT and TK are natural variants of the SW dynamics in the joint space. In the terminology of [47], they are the

Markov chains in the joint space corresponding to the two-component Gibbs sampler. The chains with transition matrices $\frac{1}{2}(K+T)$, KTK and TKT are also of interest as reversible versions of KT and TK. We show that, under SSM, all of these dynamics satisfy entropy decay with respect to ν and hence have $O(\log n)$ mixing time.

Theorem 4.5.1. *Let P be* any *of the stochastic matrices KT, TK, $\frac{1}{2}(K+T)$, KTK or TKT. SSM implies that there exists constant $\delta > 0$ such that, for all functions $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$,*

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

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

First we state the following lemma, which is proved later and will be useful in several of our proofs, including that of Theorem 4.5.1.

Lemma 4.5.2. *Let S and S' be two idempotent stochastic matrices reversible with respect to a distribution π over Γ, and let $Q = \frac{S+S'}{2}$. Suppose there exists $\delta \in (0,1)$ such that, for any positive function $f : \Gamma \mapsto \mathbb{R}$, we have $\mathrm{Ent}_\pi(Qf) \leq (1-\delta)\mathrm{Ent}_\pi(f)$. Then $\mathrm{Ent}_\pi(SS'f) \leq (1-\delta)\mathrm{Ent}_\pi(f)$ and $\mathrm{Ent}_\pi(S'Sf) \leq (1-\delta)\mathrm{Ent}_\pi(f)$.*

We are now ready to prove Theorem 4.5.1.

Proof of Theorem 4.5.1. Let us consider first the case when $P = \frac{K+T}{2}$. Since $P = P^*$, from Lemma 2.3.3 and Remark 2.3.2 it is sufficient to prove that, for all functions $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$ with $\mu[f] = 1$,

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

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

$$Pf \log(Pf) \leq \frac{1}{2}Kf \log(Kf) + \frac{1}{2}Tf \log(Tf). \tag{4.5.1}$$

If $\nu[f] = 1$, then $\nu[Pf] = \nu[Kf] = \nu[Tf] = 1$, and therefore taking expectations with respect to ν in (4.5.1) we obtain

$$\mathrm{Ent}_\nu(Pf) \leq \frac{1}{2}\left[\mathrm{Ent}_\nu(Kf) + \mathrm{Ent}_\nu(Tf)\right]. \tag{4.5.2}$$

Noting that $Kf(\sigma, A) = \nu(f \mid A)$ and $Tf(\sigma, A) = \nu(f \mid \sigma)$, the decompositions in (4.2.1) and (4.2.2) imply

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

Hence, (4.5.2) becomes

$$\mathrm{Ent}_\nu(Pf) \le \mathrm{Ent}_\nu(f) - \frac{1}{2}\nu\left[\mathrm{Ent}_\nu(f \mid A) + \mathrm{Ent}_\nu(f \mid \sigma)\right].$$

Lemma 4.3.2 now implies

$$\mathrm{Ent}_\nu(Pf) \le (1-\delta)\mathrm{Ent}_\nu(f),$$

with $\delta = 1/2C$. This proves the theorem for the case when $P = \frac{K+T}{2}$. The result for KT and TK follows from Lemma 4.5.2, by noting that $K^2 = K = K^*$ and $T^2 = T = T^*$, and noting that $(KT)^* = TK$ and $(TK)^* = KT$. (Note that in Lemma 4.5.2, we do not require the matrices to be ergodic.) Finally, the cases $P = KTK$, $P = TKT$ follow from the cases $P = KT$ and $P = TK$ with the observation that, by (4.5.5), $\mathrm{Ent}_\nu(KTKf) \le \mathrm{Ent}_\nu(TKf)$ and $\mathrm{Ent}_\nu(TKTf) \le \mathrm{Ent}_\nu(KTf)$. □

Finally, we go back and supply the missing proof of Lemma 4.5.2.

Proof of Lemma 4.5.2. Let us first show that

$$\mathrm{Ent}_\pi(Sf) \le \mathrm{Ent}_\pi(\widetilde{S}f), \tag{4.5.3}$$

where $\widetilde{S} = \frac{S+I}{2}$ is a lazy version of S; I denotes the identity matrix. To this end, define $U_n = \left[\frac{1}{2}(S+I)\right]^n$. Then we have $U_1 = \widetilde{S}$ and $U_n = \frac{(2^n-1)S+I}{2^n} \to S$ as $n \to \infty$. Therefore, (4.5.3) follows if we prove that for all $n \ge 1$

$$\mathrm{Ent}_\pi\left(U_{n+1}f\right) \le \mathrm{Ent}_\pi(U_n f). \tag{4.5.4}$$

On the other hand, if U is any stochastic matrix with stationary distribution π, then for any function $f : \Gamma \mapsto \mathbb{R}_+$ with $\pi[f] = 1$ we have $\pi[Uf] = 1$. Hence, $\mathrm{Ent}_\pi(Uf) = \pi[(Uf)\log(Uf)]$. Since U is a stochastic matrix, the convexity of the function $x \log x$ implies $(Uf)\log(Uf) \le U(f \log f)$, and so

$$\mathrm{Ent}_\pi(Uf) \le \pi[U(f\log f)] = \pi[f \log f] = \mathrm{Ent}_\pi(f). \tag{4.5.5}$$

Since $U_{n+1}f = U_1 U_n f$, applying (4.5.5) with f replaced by $U_n f$ and with $U = U_1$ proves (4.5.4) and (4.5.3). We note that since $(S')^2 = S'$,

$$\widetilde{S}S' = \frac{1}{2}(S+S')S' = QS'.$$

Applying (4.5.3) with f replaced by $S'f$ we obtain

$$\mathrm{Ent}_\pi(SS'f) \le \mathrm{Ent}_\pi(\widetilde{S}S'f) = \mathrm{Ent}_\pi(QS'f) \le (1-\delta)\mathrm{Ent}_\pi(S'f) \le (1-\delta)\mathrm{Ent}_\pi(f),$$

where the second inequality follows from the assumption that Q contracts entropy for any function and the last one follows again from (4.5.5). This completes the proof for SS'. The same argument with S and S' exchanged applies for $S'S$ and we are done. □

4.5.2 The local dynamics in the joint space

In this section, we use Lemma 4.3.2 to derive tight bounds for the local (Glauber) dynamics in the joint space; this dynamics has been recently considered in [35], but as far as we know there are no results in the literature concerning its rate of convergence to stationarity. The dynamics is defined as follows: in each step, with probability $1/2$ update a vertex and with probability $1/2$ update an edge. To update a vertex, pick $v \in V$ uniformly at random and perform a "heat-bath" update at v (i.e., replace the spin of v with a new spin sampled from the conditional distribution of the spin at v given the current spin/edge configuration); to update an edge, pick $e \in \mathbb{E}$ uniformly at random and perform a "heat-bath" update at e.

For any $v \in V$, $e \in \mathbb{E}$, let Q_v denote the stochastic matrix corresponding to the single heat-bath update at vertex v, and let W_e denote the stochastic matrix for the single heat-bath update at the edge e. Then the transition matrix P_{LOCAL} of the Glauber dynamics in the joint space is given by

$$P_{\text{LOCAL}} = \frac{1}{2|V|}\sum_{v\in V} Q_v + \frac{1}{2|\mathbb{E}|}\sum_{e\in\mathbb{E}} W_e. \tag{4.5.6}$$

Theorem 4.5.3. *SSM implies that there exists a constant $\delta > 0$ such that, for all $f : \Omega_{\text{J}} \mapsto \mathbb{R}_+$*

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

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

The mixing time bound in this theorem is asymptotically tight. This follows from the lower bounds in [77] by considering the projection of P_{LOCAL} on the spins; see Remark 4.5.4.

The heat-bath updates in the joint space are quite simple. For a vertex $v \in V$, the heat-bath update at v assigns a new spin to v chosen u.a.r. from $\{1,\dots,q\}$, provided v is isolated (i.e., there are no edges incident to v in the edge con-

figuration); otherwise, the spin at v does not change. On the other hand, the heat-bath update at $e \in \mathbb{E}$ updates the state of e only if it is monochromatic in the spin configuration; if this is the case, the new state of e corresponds to a Bernoulli(p) random variable. We note that Q_v and W_e are reversible with respect to ν. Moreover, they are projection operators in $L^2(\Omega_{\text{J}}, \nu)$; that is, $Q_v^2 = Q_v = Q_v^*$ and $W_e^2 = W_e = W_e^*$.

Proof of Theorem 4.5.3. First note that since Q_v and W_e are reversible with respect to ν, so is P_{LOCAL} and by Lemma 2.3.3 and Remark 2.3.2 it is sufficient for us to establish that

$$\text{Ent}_\nu(P_{\text{LOCAL}} f) \leq (1 - \delta/n)\text{Ent}_\nu(f) \tag{4.5.7}$$

for all functions $f : \Omega_{\text{J}} \mapsto \mathbb{R}_+$ such that $\nu[f] = 1$. Here $\delta > 0$ is a constant independent of n and the admissible boundary condition.

By the convexity of the function $x \log x$, reasoning as in (4.5.2), we can write

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

Let $\sigma_{V \setminus v}$ (resp., $A_{\mathbb{E} \setminus e}$) denote the spin (resp., edge) configuration excluding v (resp., excluding e). Since

$$Q_v f(\sigma, A) = \nu(f \,|\, \sigma_{V \setminus \{v\}}, A)$$

and

$$W_e f(\sigma, A) = \nu(f \,|\, \sigma, A_{\mathbb{E} \setminus e}),$$

from the decompositions of entropy in (4.2.1) and (4.2.2) we obtain

$$\begin{aligned} \text{Ent}_\nu(Q_v f) &= \text{Ent}_\nu(f) - \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V \setminus \{v\}}, A)\right]; \\ \text{Ent}_\nu(W_e f) &= \text{Ent}_\nu(f) \\ &\quad - \nu\left[\text{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E} \setminus e})\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Ent}_\nu(P_{\text{LOCAL}} f) \leq \text{Ent}_\nu(f) &- \frac{1}{2|V|} \sum_{v \in V} \nu\left[\text{Ent}_\nu(f \,|\, \sigma_{V \setminus \{v\}}, A)\right] \\ &- \frac{1}{2|\mathbb{E}|} \sum_{e \in \mathbb{E}} \nu\left[\text{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E} \setminus e})\right]. \end{aligned}$$

We show next that there exists a constant $C > 0$ such that

$$\mathrm{Ent}_\nu(f) \le C \sum_{v \in V} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V\setminus\{v\}}, A) \right] + C \sum_{e \in \mathbb{E}} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E}\setminus e}) \right]. \quad (4.5.8)$$

The desired estimate (4.5.7) then follows from the fact that $|\mathbb{E}| = O(|V|) = O(n)$.

To establish (4.5.8), note that by Theorem 3.3.2 we know that SSM implies approximate even/odd factorization. Then, from Lemma 4.3.5 we know that, for some constant $C_1 > 0$,

$$\mathrm{Ent}_\nu(f) \le C_1 \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_E) + \mathrm{Ent}_\nu(f \,|\, \sigma_O) \right], \quad (4.5.9)$$

where we recall that $E \subset V$ and $O \subset V$ are the even and odd sub-lattices, respectively. Since $\nu(\cdot \,|\, \sigma_O) = \otimes_{v \in E} \nu_v(\cdot \,|\, \sigma_O)$ (see (4.3.7)), the standard tensorization of entropy for product measures (see, e.g., [5]) implies

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O) \le \sum_{v \in E} \nu \left[\mathrm{Ent}_v(f \,|\, \sigma_O) \,|\, \sigma_O \right],$$

where as before we use $\mathrm{Ent}_v(\cdot \,|\, \sigma_O)$ for the entropy with respect to $\nu_v(\cdot \,|\, \sigma_O)$. From Lemma 4.3.6 we see that

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O) \le C_1 \sum_{v \in E} \nu \left[\mathrm{Ent}_v(f \,|\, \sigma_O, A) + \mathrm{Ent}_v(f \,|\, \sigma) \,|\, \sigma_O \right],$$

for some constant $C_1 > 0$.

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

$$\mathrm{Ent}_v(f \,|\, \sigma_O, A) = \mathrm{Ent}_\nu(f \,|\, \sigma_{V\setminus\{v\}}, A).$$

Let us also observe that, for every $v \in E$,

$$\mathrm{Ent}_v(f \,|\, \sigma) \le \sum_{w \in V \colon \{w,v\} \in \mathbb{E}} \mathrm{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E}\setminus\{w,v\}}). \quad (4.5.10)$$

Indeed, $\nu_v(\cdot \,|\, \sigma)$ is a product measure on $A_v = \{A_{vw},\, \{w,v\} \in \mathbb{E}\}$, and the entropy appearing on the right hand side above is simply the entropy of A_{vw} once every other spin or edge variable has been fixed. Therefore, (4.5.10) is again the standard tensorization statement for product measures. In conclusion, we have

shown that

$$\mathrm{Ent}_\nu(f \,|\, \sigma_O) \le C_1 \sum_{v \in E} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V\setminus\{v\}}, A) \,|\, \sigma_O \right] + C_1 \sum_{e \in \mathbb{E}} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E}\setminus e}) \,|\, \sigma_O \right], \tag{4.5.11}$$

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

$$\mathrm{Ent}_\nu(f \,|\, \sigma_E) \le C_1 \sum_{v \in O} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V\setminus\{v\}}, A) \,|\, \sigma_E \right] + C_1 \sum_{e \in \mathbb{E}} \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma, A_{\mathbb{E}\setminus e}) \,|\, \sigma_E \right]. \tag{4.5.12}$$

Taking expectations with respect to ν and summing (4.5.11) and (4.5.12), from (4.5.9) we obtain (4.5.8) which finishes the proof. □

Remark 4.5.4. By taking f that depends only on spins, we derive as a corollary of Theorem 4.5.3 entropy decay for the Potts model Glauber dynamics (up to a constant laziness factor to account for the probability of a site being isolated); similarly, taking f that depends only on edges, we obtain entropy decay for the corresponding Glauber dynamics for the random-cluster model. While entropy decay was previously known for the Potts Glauber dynamics under SSM [36], the same statement for the random-cluster dynamics appears to be a new result. (Note in particular that entropy decay does not follow from the mixing time results for this dynamics in [16].)

We briefly mention several other consequences of our results. First, we note that Theorem 4.5.3 can be extended to the more general case of (weighted) block dynamics for the joint space. In addition, since the "edge marginal" of the joint measure ν is the random-cluster distribution, we can show that the mixing time of the SW dynamics for the random-cluster model, which alternates between edge and joint configurations, is also $O(\log n)$ for all integer $q \ge 2$ provided SSM holds; see Section 4.7 for more details about our results for random-cluster dynamics.

Finally, we note that while our results in the joint space are all stated for the free boundary condition, they actually extend to the more general class of admissible boundary conditions; see Definition 4.3.1 in Section 4.3 for the definition of this class.

4.6 Entropy decay for the alternating scan dynamics

The fact that classical log-Sobolev inequalities do not capture the mixing time of the SW dynamics seems to be a more general phenomenon afflicting non-local

Markov chains. These chains are popular due to their presumed speed-up over Glauber dynamics and to the fact that their updates can be parallelized. With our techniques, we are able to establish entropy contraction for another standard non-local Markov chain for the Potts model known as the *alternating scan dynamics*. This chain, which is used in practice to sample from the Gibbs distribution and has received some theoretical attention [13, 114, 72], also has a "bad" log-Sobolev constant, but we can show that entropy decays at a constant rate over the steps of the chain.

In one step of the alternating scan dynamics, all the *even* vertices (i.e., those with even coordinate sum) are updated simultaneously with a new configuration distributed according to the conditional measure on the even sub-lattice given the configuration on the odd sub-lattice; the process is then repeated for the odd vertices. The key observation is that the conditional distributions on the even and odd sub-lattices are product distributions, which makes this chain particularly amenable to parallelization and thus attractive in applications.

Let P_E be the stochastic matrix corresponding to the update of the even sites conditional on the spins of the odd sites, and define P_O analogously for the odd sites. The alternating scan dynamics is the Markov chain with transition matrix $S_{EO} = P_E P_O$ (or, equivalently, $S_{OE} = P_O P_E$). Note that P_E, P_O do not commute (unless $\beta = 0$), so S_{EO} and S_{OE} are not reversible with respect to their stationary measure μ. In [13] it was shown that whenever SSM holds, the mixing time of the reversibilized version $S_{OEO} = P_O P_E P_O$ of this dynamics is $O(n)$. Here we prove a much tighter bound by showing that the alternating scan dynamics itself contracts entropy at a constant rate.

Theorem 4.6.1. *Let P be either of the stochastic matrices S_{EO} or S_{OE}. SSM implies that there exists a constant $\delta > 0$ such that, for all boundary conditions and all functions $f : \Omega \mapsto \mathbb{R}_+$,*

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

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

We note that the alternating scan dynamics is a version of so-called *systematic scan* dynamics, a variant of Glauber dynamics in which vertices are updated in some fixed, rather than random, ordering. Due to their widespread use in practice, the effect of decay of correlations properties on the speed of convergence of this class of dynamics has been widely studied; see, e.g. [53, 76, 54].

Proof of Theorem 4.6.1. We will show that the discrete entropy contraction in (2.3.5) holds for S_{EO} and S_{OE} for any positive function $f : \Omega \mapsto \mathbb{R}$ such that $\mu[f] = 1$. The

mixing time bounds then follow from Lemma 2.3.3 and the fact that $S_{EO}^* = S_{OE}$ and $S_{OE}^* = S_{EO}$. In view of Lemma 4.5.2, it is sufficient for us to establish (2.3.5) for $P := \frac{P_E+P_O}{2}$. The convexity of the function $x \log x$ implies the pointwise bound

$$(Pf)\log(Pf) \leq \frac{1}{2}(P_E f)\log(P_E f) + \frac{1}{2}(P_O f)\log(P_O f).$$

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

$$\mathrm{Ent}_\mu(Pf) = \mu[(Pf)\log(Pf)] \leq \frac{1}{2}\left[\mathrm{Ent}_\mu(P_E f) + \mathrm{Ent}_\mu(P_O f)\right]. \tag{4.6.1}$$

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

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &= \mathrm{Ent}_\mu(\mu(f \,|\, \sigma_O)) + \mu\left[\mathrm{Ent}_\mu(f \,|\, \sigma_O)\right] = \mathrm{Ent}_\mu(P_E f) + \mu\left[\mathrm{Ent}_\mu(f \,|\, \sigma_O)\right]; \\
\mathrm{Ent}_\mu(f) &= \mathrm{Ent}_\mu(\mu(f \,|\, \sigma_E)) + \mu\left[\mathrm{Ent}_\mu(f \,|\, \sigma_E)\right] = \mathrm{Ent}_\mu(P_O f) + \mu\left[\mathrm{Ent}_\mu(f \,|\, \sigma_E)\right],
\end{aligned}$$

and we see that (4.6.1) is equivalent to

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

We may now apply Theorem 3.3.2 which implies that, when SSM holds,

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

for a suitable constant $\delta \in (0,1)$. This establishes (2.3.5) for $P = \frac{P_E+P_O}{2}$. Since $P_E^2 = P_E = P_E^*$ and $P_O^2 = P_O = P_O^*$, and $(P_E P_O)^* = P_O P_E$, $(P_O P_E)^* = P_E P_O$, the remainder of the result follows from Lemma 4.5.2. □

4.7 Random-cluster dynamics

In this section we study the implications of our results for the dynamics of the *random-cluster model* for both the high and low temperatures regimes. This allows us to derive Theorem 1.3.5 from the introduction using a comparison mechanism we establish in Section 4.7.2.

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

$$\varrho(A) = \varrho_{G,p,q}(A) = \frac{1}{Z_{\mathrm{RC}}}\, p^{|A|}(1-p)^{|\mathbb{E}|-|A|} q^{c(A)}, \tag{4.7.1}$$

where $c(A)$ is the number of connected components in (V, A) and Z_{RC} is the cor-

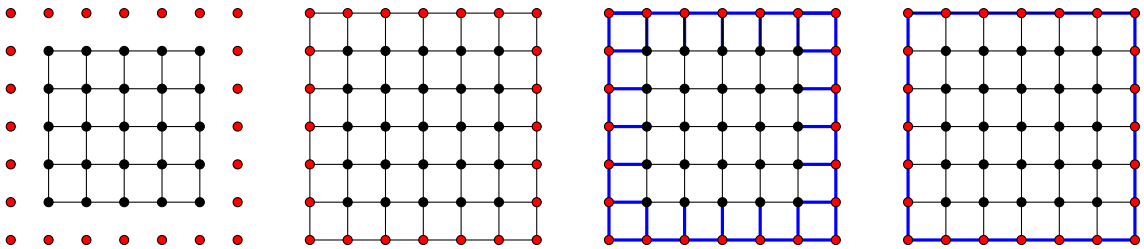

Figure 4.7.1: The figures above show four distinct admissible boundary conditions of a square region V of the joint space. The boundary condition in (a) is obtained by taking $V_0 = \partial V, \mathbb{E}_0 = \partial\mathbb{E}$, $\psi =$ "red" and $\varphi = 0$. The boundary condition in (b) is the spin-only monochromatic boundary condition obtained by taking $V_0 = \partial V$, $\mathbb{E}_0 = \emptyset$ and $\psi =$ "red"; (c) is obtained by taking $V_0 = \partial V, \mathbb{E}_0 = \partial\mathbb{E}$, $\psi =$ "red" and $\varphi = 1$ (wired edges are colored blue); note that the vertices incident to ∂V will be "red" with probability 1. Boundary condition (d) is obtained by taking $V_0 = \partial V$, $\mathbb{E}_0 = \partial\mathbb{E} \setminus \mathbb{E}_1$, $\psi =$ "red" and $\varphi = 1$. The marginal on edges of $\nu^{(\psi,\varphi)}$ in (a) is the random-cluster measure on the internal square $V \setminus \partial V$ with the free boundary condition, while in (b), (c) and (d) the edge marginal is a wired random-cluster measure over $(V, \mathbb{E})$, $(V \setminus \partial V, \mathbb{E} \setminus \partial\mathbb{E})$ and $(V, \mathbb{E} \setminus (\partial\mathbb{E} \setminus \mathbb{E}_1))$, respectively.

responding partition function. The random-cluster model was first introduced by Fortuin and Kasteleyn [62] as a unifying framework for random graphs, spin systems and electrical networks; see the book [68] for extensive background.

A *boundary condition* for the random-cluster model is a partition $\xi = \{\xi_1, \xi_2, \dots\}$ of the internal boundary ∂V of V such that all vertices in each ξ_i are constrained to be in the same connected component of any configuration A. (We can think of the vertices in ξ_i as being connected through a configuration in V^c.) These connections are considered in the counting of the connected components in (4.7.1); i.e., $c(A)$ becomes $c(A, \xi)$ (see, e.g., [15, 68]).

The distribution ϱ with a *free* boundary condition (i.e., every element of ξ is a single vertex) corresponds to the edge marginal of the joint measure also with free boundary condition (4.1.1); that is, $\varrho(A) = \sum_{\sigma: A \subseteq M(\sigma)} \nu(\sigma, A)$ and $Z_{\text{RC}} = Z_{\text{J}}$; see, e.g., [56, 68]. The *wired* boundary condition corresponds to the case when all vertices of ∂V are connected by the boundary condition (i.e., $\xi = \{\partial V\}$). More generally, if (ψ, φ) is an admissible boundary condition for the joint space (see Definition 4.3.1), we have

$$\varrho^{\psi,\varphi}(A) = \sum_{\sigma: A \subseteq M(\sigma)} \nu^{\psi,\varphi}(\sigma, A) = \frac{1}{Z^{\psi,\varphi}} p^{|A|}(1-p)^{|\mathbb{E}|-|A|} q^{c(A)-c_0(A)} \mathbf{1}(A \sim \psi)\mathbf{1}(A \sim \varphi), \tag{4.7.2}$$

where $A \sim \psi$ means that A does not connect vertices of V_0 with different colors in ψ, $A \sim \varphi$ that A and φ agree on the edges in $\mathbb{E}_0$ and $c_0(A)$ denotes the number of connected components that intersect $V_0 \subseteq \partial V$; see Figure 4.7.1 for some admissible boundary conditions.

As an example, consider the admissible boundary condition that is obtained

by taking $V_0 = \partial V$, $\mathbb{E}_0 = \partial\mathbb{E}$, with $\psi = i$ for some $i \in [q]$ (i.e., the monochromatic spin boundary condition) and $\varphi = 1$; see Figure 4.7.1(c). In this case, $\varrho^{\psi,\varphi}$ is the random-cluster measure on the cube $R = \{1, \dots, \ell - 1\}^d \subset V$ with wired boundary condition. On the other hand, the marginal on the spins is the Potts measure on R with the "all i" monochromatic boundary condition.

Another relevant random-cluster boundary condition is the one obtained by adding to the random-cluster space the edges "sticking in" from ∂V. Namely, let $\mathbb{E}_1 \subset \partial\mathbb{E}$ be the set of edges with exactly one endpoint in ∂V, and take the monochromatic boundary condition $\psi = i$ and the wired edge boundary condition on $\mathbb{E}_0 = \partial\mathbb{E} \setminus \mathbb{E}_1$. The marginal on edges is the random-cluster distribution measure on $(V, \mathbb{E} \setminus \mathbb{E}_0)$ with wired boundary condition on ∂V, while the spin marginal is the Potts measure on R with the "all i" boundary condition on ∂V; see Figure 4.7.1(d).

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

Planar duality. A useful tool in two dimensions is planar duality. Let $G_{\mathrm{d}} = (V_{\mathrm{d}}, \mathbb{E}_{\mathrm{d}})$ denote the planar dual of $G = (V, \mathbb{E})$, where $V = \{0, \dots, \ell\} \times \{0, \dots, \ell\}$ is a square region of $\mathbb{Z}^2$. That is, V_{d} corresponds to the set of faces of V, and for each $e \in \mathbb{E}$, there is a dual edge $e_{\mathrm{d}} \in \mathbb{E}_{\mathrm{d}}$ connecting the two faces bordering e. The random-cluster distribution (4.7.1) satisfies $\varrho_{G,p,q}(A) = \varrho_{G_{\mathrm{d}},p_{\mathrm{d}},q}(A_{\mathrm{d}})$, where A_{d} is the dual configuration to $A \subseteq \mathbb{E}$; i.e., $e_{\mathrm{d}} \in A_{\mathrm{d}}$ iff $e \notin A$), and

$$p_{\mathrm{d}} = \frac{q(1-p)}{q(1-p)+p}.$$

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

Since V_{d} is not a subset of $\mathbb{Z}^2$, it is convenient to consider the graph $\hat{G}_{\mathrm{d}} = (\hat{V}_{\mathrm{d}}, \hat{\mathbb{E}}_{\mathrm{d}})$ with $\hat{V}_{\mathrm{d}} = \{-1, \dots, \ell\} \times \{-1, \dots, \ell\} + (\frac{1}{2}, \frac{1}{2})$ and identify all boundary vertices of $\hat{V}_{\mathrm{d}}$ with the vertex of G_{d} corresponding to its external face. Then, $\varrho^1_{G,p,q}(A) = \varrho^0_{\hat{G}_{\mathrm{d}},p_{\mathrm{d}},q}(A_{\mathrm{d}})$ and $\varrho^0_{G,p,q}(A) = \varrho^1_{\hat{G}_{\mathrm{d}},p_{\mathrm{d}},q}(A_{\mathrm{d}})$, where the 0 and 1 superscripts denote the free and and wired boundary conditions respectively (see Section 6.1 in [68] for a detailed discussion).

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

4.7.1 SW dynamics for the random-cluster model

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

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

$\widetilde{P}_{\text{sw}}$ is reversible with respect to ϱ; see, e.g., [56, 127]. The following lemma follows from Theorem 4.5.1.

Lemma 4.7.1. *Let $\nu := \nu^{\psi,\varphi}$ be the joint distribution with an admissible boundary condition (ψ, φ). If q and $\beta = \ln(\frac{1}{1-p})$ are such that SSM holds, then the SW dynamics on random-cluster configurations with boundary conditions inherited from (ψ, φ) satisfies the discrete time entropy decay with rate δ, and its mixing time is bounded by $O(\log n)$.*

Proof. If f depends only on the edge configuration, then

$$\widetilde{P}_{\text{sw}} f(A) = \nu[\nu(f \mid \sigma) \mid A] = TKf(\sigma, A). \tag{4.7.3}$$

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

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

Theorem 4.5.1 says that, for any function f in the joint space, one has

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

In particular, for our f,

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

This is the desired discrete time entropy decay for $\widetilde{P}_{\text{sw}}$ in the edge space. □

Remark 4.7.2. The same argument in the previous proof applies to the spin dynamics. In particular, if g is a function depending only on the spin configuration,

then $P_{\text{SW}}g(\sigma) = KTg(\sigma, A)$. Repeating the previous steps with KT in place of TK one has discrete time entropy decay with rate δ for the SW dynamics on spin configurations. This provides an alternative view of the proof of Theorem 1.3.3 as a corollary of Theorem 4.5.1 for the joint space.

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

Theorem 4.7.3. *In an n-vertex square region of $\mathbb{Z}^2$ with free or wired boundary conditions, for all integer $q \geq 2$ and all $p > p_c(q)$, there exists a constant $\delta > 0$ such that for all functions $f : \{0,1\}^{\mathbb{E}} \mapsto \mathbb{R}_+$*

$$\text{Ent}_{\varrho}(\widetilde{P}_{\text{SW}} f) \leq \left(1 - \frac{\delta}{n}\right) \text{Ent}_{\varrho}(f).$$

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

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

1. choose an edge $e \in \mathbb{E}$ uniformly at random;

2. let $A_{t+1} = A_t \cup \{e\}$ with probability

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

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

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

Theorem 4.7.3 follows from the following two results.

Lemma 4.7.4. *There exists a constant $c > 0$ such that, for every function $f : \{0,1\}^{\mathbb{E}} \mapsto \mathbb{R}$,*

$$\mathcal{D}_{\widetilde{P}_{\text{SW}}}(f, f) \geq c \cdot \mathcal{D}_{P_{\text{HB}}}(f, f).$$

Lemma 4.7.5. *For all integer $q \geq 2$ and all $p > p_c(q)$, there exists a constant $\delta > 0$ such that, for every function $f : \{0,1\}^{\mathbb{E}} \mapsto \mathbb{R}_+$,*

$$\mathcal{D}_{P_{\text{HB}}}(\sqrt{f}, \sqrt{f}) \geq \frac{\delta}{n} \cdot \text{Ent}_{\varrho}(f).$$

Proof of Theorem 4.7.3. Lemmas 4.7.4 and 4.7.5 imply

$$\mathcal{D}_{\widetilde{P}_{\text{SW}}}(\sqrt{f}, \sqrt{f}) \geq \frac{c\delta}{n} \cdot \text{Ent}_{\varrho}(f). \tag{4.7.4}$$

In words, this says that the SW dynamics on random-cluster configurations when $p > p_c(q)$ satisfies a standard log-Sobolev inequality with constant $\frac{c\delta}{n}$. An inequality of Miclo relating the standard log-Sobolev inequality and discrete time entropy decay (see Proposition 6 in [104]) shows that (4.7.4) implies the entropy decay bound

$$\text{Ent}_{\varrho}(\widetilde{P}_{\text{SW}} f) \leq \left(1 - \frac{\delta c}{n}\right) \text{Ent}_{\varrho}(f),$$

and the mixing time bound follows from Lemma 2.3.3 and Remark 2.3.2 since $\widetilde{P}_{\text{SW}} = \widetilde{P}^*_{\text{SW}}$. □

It remains to prove Lemmas 4.7.4 and 4.7.5. We note that a version of the comparison inequality in Lemma 4.7.4 was proved in [127] (see Theorem 4.8 there), but it is stated for the spectral gap under the free boundary condition.

In both of these proofs, we consider the single-bond variant of the Glauber dynamics. In one step of this chain every connected component is assigned a spin from $[q]$ uniformly at random; a random edge e is then chosen and if the endpoints of e are monochromatic, then the edge is added to the configuration with probability p and deleted otherwise. The state of e does not change if its endpoints are bi-chromatic. Note that this chain is the projection onto edges of the local dynamics on the joint space, see (4.5.6); in particular, the update at the edge e corresponds to W_e. Let P_{SB} denote the transition matrix of the single bond dynamics, which is reversible with respect to ϱ. The Dirichlet form associated to this chain satisfies

$$\mathcal{D}_{P_{\text{SB}}}(f,f) = \langle (I - P_{\text{SB}})f, f\rangle_{\varrho} = \varrho\left[((I - P_{\text{SB}})f) \cdot f\right] = \frac{1}{|\mathbb{E}|}\sum_{e\in\mathbb{E}} \nu\left[\text{Var}_{\nu}(f \,|\, \sigma, A_{\mathbb{E}\setminus e})\right] \tag{4.7.5}$$

since

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

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

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

$$c_1 P_{\text{SB}}(A,B) \leq P_{\text{HB}}(A,B) \leq c_2 P_{\text{SB}}(A,B)$$

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

$$c_1 \mathcal{D}_{P_{\text{SB}}}(f,f) \leq \mathcal{D}_{P_{\text{HB}}}(f,f) \leq c_2 \mathcal{D}_{P_{\text{SB}}}(f,f), \tag{4.7.6}$$

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

Proof of Lemma 4.7.4. The Dirichlet form associated with $\widetilde{P}_{\text{SW}}$ is given by

$$\mathcal{D}_{\widetilde{P}_{\text{SW}}}(f,g) = \langle (I - \widetilde{P}_{\text{SW}})f, g \rangle_{\varrho} = \varrho \left[((I - \widetilde{P}_{\text{SW}})f) \cdot g \right],$$

and since $\widetilde{P}_{\text{SW}} f(A) = \nu[\nu[f \,|\, \sigma] \,|\, A]$, we obtain

$$\mathcal{D}_{\widetilde{P}_{\text{SW}}}(f,f) = \nu\left[(f - \nu[\nu[f \,|\, \sigma] \,|\, A]) \cdot f\right] = \nu\left[(f - \nu[f \,|\, \sigma]) \cdot f\right] = \nu\left[\text{Var}_\nu(f \,|\, \sigma)\right].$$

Then, for any function $f \geq 0$,

$$\begin{aligned}
\mathcal{D}_{\widetilde{P}_{\text{SW}}}(\sqrt{f}, \sqrt{f}) &= \nu\left[\text{Var}_\nu(\sqrt{f} \,|\, \sigma)\right] \\
&\geq \frac{1}{|\mathbb{E}|} \sum_{e \in \mathbb{E}} \nu\left[\text{Var}_\nu(\sqrt{f} \,|\, \sigma, A_{\mathbb{E}\setminus e})\right] \\
&= \mathcal{D}_{P_{\text{SB}}}(\sqrt{f}, \sqrt{f}),
\end{aligned}$$

where we have used (4.7.5) and the fact that, for any $e \in \mathbb{E}$,

$$\nu\left[\text{Var}_\nu(\sqrt{f} \,|\, \sigma)\right] \geq \nu\left[\text{Var}_\nu(\sqrt{f} \,|\, \sigma, A_{\mathbb{E}\setminus e})\right]$$

by monotonicity of the variance functional. The result then follows from (4.7.6). □

Proof of Lemma 4.7.5. By duality (see discussion at the beginning of the section), we have

$$\mathcal{D}_{P_{\text{HB}}}(\sqrt{f}, \sqrt{f}) = \mathcal{D}_{P^{\text{d}}_{\text{HB}}}(\sqrt{f_{\text{d}}}, \sqrt{f_{\text{d}}}), \tag{4.7.7}$$

where f_{d} is the function such that $f_{\text{d}}(A_{\text{d}}) = f(A)$ and P^{d}_{HB} is the transition matrix corresponding to the dual of ϱ.

Thus, if $\mathcal{D}_{P_{\text{HB}}}$ is at low temperature ($p > p_c(q)$), then $\mathcal{D}_{P^{\text{d}}_{\text{HB}}}$ is at high temperature

($p < p_c(q)$). Moreover, from (4.7.5) and (4.7.6),

$$\mathcal{D}_{P^{\mathrm{d}}_{\mathrm{HB}}}(\sqrt{f},\sqrt{f}) \geq c_1 \mathcal{D}_{P^{\mathrm{d}}_{\mathrm{SB}}}(\sqrt{f},\sqrt{f}) = \frac{c_1}{|\mathbb{E}|}\sum_{e\in\mathbb{E}} \nu_{\mathrm{d}}\left[\mathrm{Var}_{\nu_{\mathrm{d}}}(\sqrt{f}\,|\,\sigma, A_{\mathbb{E}\setminus e})\right],$$

where ν_{d} is the dual joint measure. Specifically, if ϱ_{d} is the dual measure of ϱ (and the stationary distribution of $P^{\mathrm{d}}_{\mathrm{HB}}$), ν_{d} is a joint measure whose edge marginal is ϱ_d. Observe that since ϱ is a random-cluster distribution on the square region $V = \{0,\dots,\ell\}\times\{0,\dots,\ell\}$ with free (or wired) boundary condition, ϱ_d is a distribution over $\hat{V}_{\mathrm{d}} = \{-1,\dots,\ell\}\times\{-1,\dots,\ell\} + (\frac{1}{2},\frac{1}{2})$ with wired (or free) boundary condition. As discussed earlier, in either case there exists a joint measure with an admissible boundary condition whose edge marginal is ϱ_d.

Observe also that, as before, with a slight abuse of notation, we also use f for the "lift" of f to the joint space. Now, as in (4.3.10) we know that for some constant $C = C(p,q)$, for all $e\in\mathbb{E}$ and for all $f\geq 0$,

$$\mathrm{Var}_{\nu_{\mathrm{d}}}(\sqrt{f}\,|\,\sigma, A_{\mathbb{E}\setminus e}) \geq C^{-1}\mathrm{Ent}_{\nu_{\mathrm{d}}}(f\,|\,\sigma, A_{\mathbb{E}\setminus e}).$$

Therefore,

$$\mathcal{D}_{P^{\mathrm{d}}_{\mathrm{HB}}}(\sqrt{f},\sqrt{f}) \geq \frac{c_1 C^{-1}}{|\mathbb{E}|}\sum_{e\in\mathbb{E}} \nu_{\mathrm{d}}\left[\mathrm{Ent}_{\nu_{\mathrm{d}}}(f\,|\,\sigma, A_{\mathbb{E}\setminus e})\right].$$

Since for $p < p_c(q)$ and $q\geq 2$ the SSM property holds, we can use (4.5.8) to obtain

$$\sum_{e\in\mathbb{E}} \nu_{\mathrm{d}}\left[\mathrm{Ent}_{\nu_{\mathrm{d}}}(f\,|\,\sigma, A_{\mathbb{E}\setminus e})\right] \geq \delta_1 \mathrm{Ent}_{\nu_{\mathrm{d}}}(f).$$

Indeed, if f is a function of edges only then the first term on the right hand side of (4.5.8) is zero. Moreover for such an f we have $\mathrm{Ent}_{\nu_{\mathrm{d}}}(f) = \mathrm{Ent}_{\varrho_{\mathrm{d}}}(f)$. Summarizing, we have proved, for all $f\geq 0$,

$$\mathcal{D}_{P^{\mathrm{d}}_{\mathrm{HB}}}(\sqrt{f_{\mathrm{d}}},\sqrt{f_{\mathrm{d}}}) \geq \frac{\delta_2}{n}\,\mathrm{Ent}_{\varrho_{\mathrm{d}}}(f_{\mathrm{d}}), \tag{4.7.8}$$

for a suitable constant $\delta_2 > 0$. The result follows from (4.7.7) and the fact that $\mathrm{Ent}_{\varrho_{\mathrm{d}}}(f_{\mathrm{d}}) = \mathrm{Ent}_{\varrho}(f)$. □

Remark 4.7.6. We remark that (4.7.8) says that the heat-bath Glauber dynamics for the random-cluster model in square regions of $\mathbb{Z}^2$ with free or wired boundary conditions satisfies the standard log-Sobolev inequality with constant δ/n for some $\delta = \delta(p,q)$ for all $p \neq p_c(q)$. This bound is optimal up to a multiplicative constant, as can be seen by choosing an appropriate test function.

4.7.2 Decay for spins from decay for edges and vice versa

We will use Theorem 4.7.3 to deduce our low temperature results for the SW dynamics on spin configurations. We do so using the following entropy contraction "transfer" result between the spin and edge variants of the SW dynamics. A similar comparison result for the spectral gap was provided by Ullrich [127].

Lemma 4.7.7. *Suppose we know that the SW dynamics on edges with invariant measure ϱ, corresponding to an n-vertex square region V with some boundary condition, has entropy decay with rate δ. Then the SW dynamics on spins on V, with any boundary condition inherited from a joint measure ν whose marginal on edges equals ϱ, satisfies the same entropy decay (asymptotically) and has the same mixing time bound $T_{\mathrm{mix}} = O(\delta^{-1} \log n)$. The same applies with the roles of spins and edges reversed.*

Proof. The assumption on ϱ says that

$$\mathrm{Ent}_{\varrho}(\widetilde{P}_{SW} g) \leq (1-\delta)\mathrm{Ent}_{\varrho}(g), \tag{4.7.9}$$

for any function $g = g(A)$, $A \subset \mathbb{E}$. Recalling (4.7.3) we see that (4.7.9) can be rewritten as

$$\mathrm{Ent}_{\nu}(TKg) \leq (1-\delta)\mathrm{Ent}_{\nu}(g),$$

for any $g = g(A)$ and any joint measure ν such that the marginal on edges equals ϱ. Now, let $f = f(\sigma)$ be any function depending only on the spin configuration. Since $g = Tf$ depends only on the edge configuration, we have

$$\mathrm{Ent}_{\nu}(TKTf) \leq (1-\delta)\mathrm{Ent}_{\nu}(Tf). \tag{4.7.10}$$

If we apply (4.7.10) with f replaced by $(KT)^{\ell-1} f$, then

$$\mathrm{Ent}_{\nu}(T(KT)^{\ell} f) \leq (1-\delta)\mathrm{Ent}_{\nu}(T(KT)^{\ell-1} f),$$

for any $\ell \in \mathbb{N}$. Iterating this inequality we find, for any $\ell \in \mathbb{N}$,

$$\mathrm{Ent}_{\nu}(T(KT)^{\ell} f) \leq (1-\delta)^{\ell}\mathrm{Ent}_{\nu}(Tf). \tag{4.7.11}$$

Recalling that $P_{SW}^{\ell} f = (KT)^{\ell} f$, from (4.7.11) we get

$$\begin{aligned}
\mathrm{Ent}_{\mu}(P_{SW}^{\ell} f) &= \mathrm{Ent}_{\nu}((KT)^{\ell} f) \\
&= \mathrm{Ent}_{\nu}(KT(KT)^{\ell-1} f) \\
&\leq \mathrm{Ent}_{\nu}(T(KT)^{\ell-1} f) \\
&\leq (1-\delta)^{\ell-1} \mathrm{Ent}_{\nu}(Tf) \\
&\leq (1-\delta)^{\ell-1} \mathrm{Ent}_{\nu}(f) = (1-\delta)^{\ell-1} \mathrm{Ent}_{\mu}(f),
\end{aligned}$$

where the first inequality follows from (4.5.5). This shows that the discrete time entropy decay for SW on spins is asymptotically the same as the one assumed for SW on edges, and Lemma 2.3.3 allows us to conclude the desired mixing time bound. The same argument (with KT replaced by TK) shows that if we assume an entropy decay for spins then we obtain (asymptotically) the same entropy decay for edges, and therefore the same mixing time bound. □

We can now provide the proof of Theorem 1.3.5 from the introduction.

Proof of Theorem 1.3.5. From the discussion at the beginning of Section 4.7, note that there is an admissible boundary condition in the joint space for which the edge marginal is the random-cluster measure on a square region of $\mathbb{Z}^2$ with a wired boundary condition, and the spin marginal is the monochromatic boundary condition. The result then follows from Theorem 4.7.3 and Lemma 4.7.7. □

Chapter 5

Spectral independence and q-spin systems

5.1 Proof overviews for main results

In this chapter, we consider arbitrary $q-$spin systems on $G = (V, E)$ where the maximum degree of the graph G is independent on $|V|$, and we sketch the proof of our key technical results showed in section 1.4 of the introduction. We begin in Section 5.1.1 with an overview of our proof that spectral independence implies optimal bounds for arbitrary block dynamics and the Swendsen-Wang dynamics (namely, Theorems 1.4.1 and 1.4.2). In Section 5.1.2 we highlight the proofs of Theorems 1.4.3, and 1.4.4 that a contractive coupling for an arbitrary local dynamics implies spectral independence.

5.1.1 Optimal mixing under spectral independence

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 that to establish optimal mixing for an arbitrary choice of block dynamics it suffices to prove general block factorization (GBF); see Lemma 2.4.6 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 [42] and Theorem 5.4.1. Note, ℓ-UBF refers to the block factorization where the weights α are uniform over all subsets of size ℓ; see Definition 5.1.1 below. The key step in the proof of Theorem 1.4.1 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 $\ell \leq n$, let

$\binom{V}{\ell}$ denote the collection of all subsets of V of size ℓ.

Definition 5.1.1 (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 $f : \Omega \to \mathbb{R}_+$

$$\frac{\ell}{n} \operatorname{Ent}(f) \leq C_{\text{UBF}} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\operatorname{Ent}_S(f)]. \tag{5.1.1}$$

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

Theorem 5.1.2. *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 $0 < \theta \leq \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, then GBF holds with constant $C_{\text{GBF}} = C_{\text{UBF}} \times O\left((\theta\, b^2)^{-1} \log(1/b)\Delta^3\right)$.*

This and the already known $\lceil \theta n \rceil$-UBF implies Theorem 1.4.1 from the introduction:

Proof of Theorem 1.4.1. For a spin system that is η-spectrally independent and b-marginally bounded, $\lceil \theta n \rceil$-UBF holds with constant $C_{\text{UBF}} = (\frac{1}{\theta})^{O(\frac{\eta}{b})}$ (see Theorem 5.4.1). Then, taking $\theta = \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, Theorem 5.1.2 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 1.4.1 is Theorem 5.1.2. To establish Theorem 5.1.2 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 $k \leq \Delta + 1$. Let $\{V_1, ..., V_k\}$ denote the independent sets $V_i \subset V$ corresponding to a k-partition of G. Theorem 5.1.2 follows immediately from the following factorization statements.

Lemma 5.1.3. *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 $\theta \leq \frac{b^{2(\Delta+1)}}{4e\Delta^2}$. Then,*

$$\operatorname{Ent}(f) \leq K C_{\text{UBF}} \sum_{i=1}^{k} \mu[\operatorname{Ent}_{V_i}(f)], \tag{5.1.2}$$

where the constant K satisfies $K = O(\Delta^2(\theta\, b^2)^{-1} \log(1/b))$. We refer to inequality (5.1.2) *as a k-partite factorization of entropy with constant KC_{UBF}.*

Lemma 5.1.4. *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 5.2. The main idea behind the proof of Lemma 5.1.3 can be roughly explained as follows. If the sets S in (5.1.1) were all independent sets, then a suitable decomposition of the entropy functional would imply the desired conclusion. Using a tensorization argument from [27], 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 [27] and [42] together with a new conditioning argument. The proof of Lemma 5.1.4 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, see lemma 2.4.4. Lemma 5.1.4 also generalizes proposition 3.3.1 in chapter 3.

Finally, to prove Theorem 1.4.2, that is the optimal mixing results for the SW dynamics, our strategy is based on establishing the *spin/edge factorization* of entropy. In chapter 4 we showed that the spin/edge factorization of entropy implies $O(\log n)$ mixing of the SW dynamics on any graph, see lemma 4.1.4. To prove Theorem 1.4.2, we show that k-partite factorization of entropy for μ implies spin/edge factorization of entropy. This requires a nontrivial adaptation of Lemma 4.1.3 established in chapter 4 in the special case of bipartite graphs. The proof of Theorem 1.4.2 is provided in Section 5.5.

5.1.2 Spectral independence via contractivity

Here we outline our proofs of Theorems 1.4.3 and 1.4.4. 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 5.3.1 for relevant definitions. The im-

portant 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 [73, 117] to bound the Wasserstein distance of two distributions; see Lemma 5.3.4. 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 [73, 117] were specialized for the binary product space and the Glauber dynamics. Here, we establish our Lemma 5.3.4 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 in the article [88] have been established similar results to Theorem 1.4.4, which also implied Theorem 1.4.8 for the Glauber dynamics. In [88] the author concluded a version of Theorem 1.4.6 as well, but required the stronger assumption that the row sum of the Dobrushin dependency matrix is bounded. Using our Theorem 1.4.3 we only require a bound on the spectral radius which is the weakest assumption of this type; see Remark 5.3.5.

We now turn to the organization of this chapter. In Section 5.2 we prove that uniform block factorization implies general block factorization of entropy. In Section 5.3, we establish spectral independence if the distribution admits a contractive Markov chain. In Section 5.4, we reformulate the result of [42] 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 5.4 that spectral independence implies approximate subadditivity of entropy, see Theorem 5.4.1. Finally, in Section 5.5 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.

5.2 Uniform block factorization implies general block factorization

We provide in this section the proofs of Lemmas 5.1.4 and 5.1.3. Recall that these are the key ingredients for proving Theorem 5.1.2.

Proof of Lemma 5.1.4. 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 $\tau \in \Omega_{V \setminus V_j}$, $\mu^{\tau}_{V_j}$ is a product mea-

sure on $\Omega^\tau_{V_j}$. Therefore, we can apply Lemma 2.4.4 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

$$\delta(\hat{\alpha})\,\mathrm{Ent}^\tau_{V_j}(f) \le \sum_{U\subset V_j}\hat{\alpha}_U\,\mu^\tau_{V_j}[\mathrm{Ent}_U(f)] = \omega^{-1}\sum_{B\subset V}\alpha_B\,\mu^\tau_{V_j}[\mathrm{Ent}_{V_j\cap B}(f)]. \quad (5.2.1)$$

Observe that

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

and from (2.4.3) we have $\mu[\mathrm{Ent}_{V_j\cap B}(f)] \le \mu[\mathrm{Ent}_B(f)]$. Hence, taking expectation in (5.2.1) with respect to μ we obtain

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

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

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

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

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

Hence, GBF holds with constant Ck. □

Proof of Lemma 5.1.3. Since $\lceil\theta n\rceil$-UBF holds by assumption, setting $C = C_{\text{UBF}}$ one has

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

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 2.4.1 and 2.4.5, one has the decomposition

$$\mu\left[\mathrm{Ent}_S(f)\right] = \sum_{i\ge1}\mu\left[\mathrm{Ent}_{S_{<i+1}}(\mu_{S_{<i}}(f))\right] = \sum_{i\ge1}\mu[\mathrm{Ent}_{S_i}(\mu_{S_{<i}}(f))], \quad (5.2.2)$$

where we have used Eq. (2.4.6) 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 $j = 1, ..., k$ one has

$$\mu\left[\mathrm{Ent}_{V_j \cap S_i}(\mu_{S_{<i}}(f))\right] \leq \mu\left[\mathrm{Ent}_{V_j \cap S_i}(\mu_{V_j \cap S_{<i}}(f))\right]. \tag{5.2.3}$$

To see this, we apply Lemma 2.4.5 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 (5.2.3) from Eq. (2.4.7).

Summarizing, we have obtained

$$\mathrm{Ent}(f) \leq \frac{C}{\theta} \sum_{j=1}^{k} \mathbb{E}\left[\sum_{i \geq 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j \cap S_i}(\mu_{V_j \cap S_{<i}}(f))\right]\right]. \tag{5.2.4}$$

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

$$\mathbb{E}\left[\sum_{i \geq 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j \cap S_i}(\mu_{V_j \cap S_{<i}}(f))\right]\right] \leq C' \mu\left[\mathrm{Ent}_{V_j}(f)\right], \tag{5.2.5}$$

with $C' = O\left(\frac{\log(1/b)}{b^2}\Delta^2\right)$. Combined with (5.2.4), this concludes the proof of the lemma.

Let us fix j and let $v_1, v_2, \ldots$ 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 $i \geq 1$, $\mu_{V_j \cap S_i}$ is a product measure, Lemmas 2.4.1 and 2.4.5 (as in (5.2.2)) 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 $v_h, v_{h+1}, \ldots, v_{|V_j \cap S|}$.

Using this decomposition and rearranging one finds

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

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 2.4.1 and 2.4.5 that

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

the inequality follows from (2.4.3).

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, (5.2.5) follows if we prove that

$$\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[\Gamma(S(v))\,|\,V_j\cap S=W\right]\leq C'. \tag{5.2.6}$$

Now, for a b marginally bounded spin system, it follows from Lemma 4.2 in [42] and (2.4.3) that

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

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

$$\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[\Gamma(S(v))\,|\,V_j\cap S=W\right]\leq\zeta\cdot\max_{W\subset V_j}\max_{v\in W}\mathbb{E}\left[|S(v)|^3z^{|S(v)|}\;\middle|\;V_j\cap S=W\right]. \tag{5.2.7}$$

To bound the expectation on the right-hand-side of (5.2.7), 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 $|U|\geq a$, $v\in U$ 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 $\Delta|S(v) \cap V_{\neq j}| \geq |S(v) \cap V_j|$, and so

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

which implies that $|S_2| = |S(v) \cap V_{\neq j}| \geq a/(\Delta + 1)$. 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 $a \geq 1$ we get

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

To estimate the size of the connected component $T_2(v)$ one can use Lemma 4.3 from [42], which implies that for any integer $m \geq 1$,

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

Indeed, the only difference with respect to Lemma 4.3 from [42] is that we have maximum degree Δ^2 here instead of Δ. In particular, if $2e\Delta^2\theta \leq 1/2$, using $\frac{\ell}{n} \leq 2\theta$,

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

It follows that

$$\begin{aligned}\mathbb{E}\left[|S(v)|^3 z^{|S(v)|} \,|\, V_j \cap S = W\right] &= \sum_{a\geq 1} a^3 z^a \cdot \mathbb{P}\left(|S(v)| = a \,|\, V_j \cap S = W\right) \\ &\leq \Delta^{-2} \sum_{a\geq 1} a^3 (2e\Delta^2\theta z^{\Delta+1})^{\lfloor\frac{a}{\Delta+1}\rfloor} \leq C_1\Delta^2,\end{aligned}$$

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

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

Hence, (5.2.6) and (5.2.5) hold with $C' = C_1\zeta\Delta^2$, and consequently k-partite factorization holds with constant $C_{\text{UBF}}C_1\zeta\Delta^2/\theta$. □

5.3 A contractive distribution is spectrally independent

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 1.4.3 concerned with Glauber dynamics and Hamming metric in Section 5.3.1; this will serve as a concrete example to illustrate our approach for establishing spectral independence. In Section 5.3.2, we consider arbitrary metric and prove Theorem 1.4.3. Finally, we consider general Markov chains and metrics in Section 5.3.3 and prove Theorem 1.4.4.

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

In this section, we prove a simpler version of Theorem 1.4.3, which already gives the main idea of our proof approach for establishing spectral independence.

We first give the formal definition of κ-contraction of the distribution μ.

Definition 5.3.1. Let $\mathcal{P} = \{P^\tau : \tau \in \mathcal{T}\}$ denote a collection of Markov chains associated with μ where each P^τ is a Markov chain with stationary distribution μ^τ 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:

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

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 5.3.2. *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 $\kappa \leq 1 - \epsilon/n$, then $\eta \leq 2/\epsilon$.*

Remark 5.3.3. We define the Glauber dynamics P^τ_{GL} for the conditional distribution μ^τ with a pinning τ on $U \subset V$ 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 5.3.2 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 5.3.2, the contraction rate satisfies $\kappa_{\ell} \leq 1 - \epsilon/\ell$, so we eventually get $\eta \leq 2/\epsilon$ just as from Theorem 5.3.2.

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 $\lambda_1(J^{\tau}) \leq \eta$ 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 [4, 41, 40, 132, 42] is to upper bound the infinity norm of J. More specifically, for each $(x, a) \in \mathcal{X}$ we define

$$S(x, a) = \sum_{(y,a') \in \mathcal{X}} |J(x, a; y, a')| \tag{5.3.1}$$

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 $S(x, a) \leq \eta$ for all $(x, a) \in \mathcal{X}$, then it immediately follows that

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

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

$$f(\sigma) = \sum_{(y,a') \in \mathcal{X}} t(x, a; y, a') \, \mathbf{1}_{\{\sigma_y = a'\}}, \tag{5.3.2}$$

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

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

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., $L_d(f) = \inf\{L \geq 0 : f \text{ is } L\text{-Lipschitz w.r.t. } d\}$. 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 [131],

$$W_{1,d}(\mu,\nu) = \sup_{\substack{f:\Omega\to\mathbb{R}\\ L_d(f)\le 1}} \mathbb{E}_\mu f - \mathbb{E}_\nu f. \tag{5.3.3}$$

Observe that, the function f defined by (5.3.2) 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 (5.3.3) 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 [73, 117], 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 [73, Theorem 3.1] and [117, 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 5.3.3.

Lemma 5.3.4. *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 5.3.4 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 5.3.4. The proof imitates the arguments from [73, 117]. 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,

$$h = \sum_{t=0}^{\infty} P^t \bar{f}. \tag{5.3.4}$$

See Lemma 2.1 in [73] 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 (5.3.3) 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 (5.3.3). Since μ is κ-contractive w.r.t. P and d, for all $\sigma, \tau \in \Omega$ and every integer $t \ge 1$ we have

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

We then deduce that

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

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

Next, we show how to remove the assumption that P is irreducible. Observe that in the proof above we only need the irreducibility of P to guarantee that the function h given by (5.3.4) is well-defined; i.e., the series on the right-hand side of (5.3.4) is convergent. The rest of the proof does not require the irreducibility of P. In fact, one can deduce the convergence of (5.3.4) solely from the contraction

of P. Note that for all $\sigma \in \Omega$,

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

where the first equality follows from $\mathbb{E}_\mu P^t\bar{f} = \mathbb{E}_\mu\bar{f} = 0$. Since Ω is finite, to show that (5.3.4) 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

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

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

Given Lemma 5.3.4, we can now complete the proof of Theorem 5.3.2.

Proof of Theorem 5.3.2. For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 5.3.4 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le \frac{L_{d_{\mathrm{H}}}(f)}{1-\kappa}\,\mathbb{E}_\nu\left[W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot),Q(\sigma,\cdot))\right] \tag{5.3.5}$$

where $S(x,a)$ is given by (5.3.1), f is given by (5.3.2), 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 $\sigma \in \Omega^{(x,a)}$,

$$W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot),Q(\sigma,\cdot)) \le \frac{1}{n}. \tag{5.3.6}$$

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 $d_{\mathrm{H}}(\sigma_1,\sigma_2) \le 1$ where the discrepancy is caused by the pinning $\sigma_x = a$. Therefore, the 1-Wasserstein distance between σ_1 and σ_2 is upper bounded by $1/n$; this justifies our claim. Combining $L_{d_{\mathrm{H}}}(f) \le 2$ and (5.3.6), we obtain from (5.3.5) that $S(x,a) \le \frac{2}{(1-\kappa)n}$ for each (x,a); consequently, $\lambda_1(J) \le \frac{2}{(1-\kappa)n}$. The same argument holds for μ^τ under any pinning τ as well, and spectral independence then

follows. □

5.3.2 Contraction for Glauber dynamics and general metrics

In this section, we generalize the Hamming metric assumption in Theorem 5.3.2 to any weighted Hamming metric or any metric equivalent to Hamming, which establishes Theorem 1.4.3. We restate it here for convenience.

Theorem 1.4.3.

(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 $\eta \leq \frac{2\gamma^2}{\epsilon}$.

We prove the two cases of Theorem 1.4.3 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 5.3.2 does not work directly in this scenario. The reason is that the right-hand side of (5.3.5), 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 1.4.3(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

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

Such weighted sums were considered in [41, Lemma 22] to deduce spectral independence. We claim that if $S_w(x, a) \leq \eta \, w(x)$ for all $(x, a) \in \mathcal{X}$ for some $\eta > 0$, then $\lambda_1(J) \leq \eta$. 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 $\|W^{-1} J W\|_\infty \leq \eta$ and

thus $\lambda_1(J) = \lambda_1(W^{-1}JW) \leq \eta$. 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

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

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

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

It then follows from Lemma 5.3.4 that

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

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

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

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 $S_w(x,a) \leq \frac{2w(x)}{(1-\kappa)n}$ for every $(x,a) \in \mathcal{X}$, implying that $\lambda_1(J) \leq \frac{2}{(1-\kappa)n}$. 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 1.4.3. Recall that a metric d on Ω is said to be γ-equivalent (to the Hamming metric) for some $\gamma > 1$ if for all $\sigma, \tau \in \Omega$

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

To prove the second part, we follow the proof approach for Theorem 5.3.2, and in particular the right-hand side of (5.3.7) below (analogous to (5.3.5)) can be upper bounded using the γ-equivalence.

Proof of Theorem 1.4.3(2). For every $(x,a) \in \mathcal{X}$, we deduce from Lemma 5.3.4 that

$$S(x,a) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \leq \frac{L_d(f)}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \right] \tag{5.3.7}$$

where $S(x,a)$ and f are defined by (5.3.1), (5.3.2) 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

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

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

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

where the last inequality is (5.3.6). Thus, we obtain from (5.3.7) that $S(x,a) \leq \frac{2\gamma^2}{(1-\kappa)n}$. The rest of the proof is the same as Theorem 5.3.2. □

Application: Dobrushin uniqueness condition

As an application of Theorem 1.4.3, we show that the Dobrushin uniqueness condition, as well as its generalizations [76, 54], implies spectral independence. Recall that the Dobrushin dependency matrix R is a $|V| \times |V|$ matrix defined as $R(x,x) = 0$ and

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

where $\mathcal{S}_{x,y}$ is the set of pairs of configurations on $V \setminus \{y\}$ that differ at most at x. Denote the spectral radius of a square matrix M by $\varrho(M)$. If M is nonnegative, then $\varrho(M)$ is an eigenvalue of M by the Perron-Frobenius theorem. We prove Theorem 1.4.6 from the introduction.

Theorem 1.4.6. *If the Dobrushin dependency matrix R satisfies $\varrho(R) \leq 1 - \epsilon$ for some $\epsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\epsilon$.*

Remark 5.3.5. If $\|R\|_\infty < 1$, then the Glauber dynamics mixes rapidly by a simple application of the path coupling method of Bubley and Dyer [21]. The same is true under the Dobrushin uniqueness condition, i.e., when $\|R\|_1 < 1$. Hayes [76] generalized the condition to the spectral norm $\|R\|_2 < 1$. Dyer, Goldberg, and Jerrum [54] 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 1.4.6 is technically better than previous works since for a nonnegative matrix R one has $\varrho(R) \leq \|R\|$ for any matrix norm, and the inequality can be strict for all norms when R is not irreducible; see [54] 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 5.3.6 ([54, Lemma 20]). *If $w \in \mathbb{R}_+^V$ is a positive vector such that $Rw \leq (1-\epsilon)w$ entrywisely, then μ is $(1-\epsilon/n)$-contractive w.r.t. the Glauber dynamics and the w-weighted Hamming metric $d = d_w$.*

The following fact about nonnegative matrices is helpful.

Lemma 5.3.7 ([103, Example 7.10.2]). *If $M, N \in \mathbb{R}_+^{n\times n}$ are two nonnegative square matrices such that $M \leq N$ entrywisely, then $\varrho(M) \leq \varrho(N)$.*

We give below the proof of Theorem 1.4.6.

Proof of Theorem 1.4.6. Consider first the case that there is no pinning. If the Dobrushin dependency matrix R is irreducible, then the right principal eigenvector w associated with the eigenvalue $\varrho(R)$ satisfies $Rw = \varrho(R)w \leq (1-\epsilon)w$ and $w > 0$ by the Perron-Frobenius theorem. Hence, Lemma 5.3.6 and (the proof of) Theorem 1.4.3(1) immediately yield $\lambda_1(J) \leq 2/\epsilon$. However, if R is reducible, we cannot use the principal eigenvector directly since it may have zero entries. We instead consider the matrix $R_\delta = R + \delta O$ where O is the all-one matrix and $\delta > 0$ is a tiny constant. Let w_δ be the right principal eigenvector of R_δ associated with the eigenvalue $\varrho(R_\delta)$. Since R_δ is irreducible, $w_\delta > 0$ by the Perron-Frobenius theorem. Moreover, $Rw_\delta \leq R_\delta w_\delta = \varrho(R_\delta)w_\delta$. Since $\lim_{\delta\to 0} R_\delta = R$, we have $\lim_{\delta\to 0}\varrho(R_\delta) = \varrho(R)$; see, e.g., Remark 3.4 in [2]. Thus, $\varrho(R_\delta) < 1$ for sufficiently small δ. Then by Lemma 5.3.6 and Theorem 1.4.3(1), for δ small enough, we have $\lambda_1(J) \leq 2/(1-\varrho(R_\delta))$. Taking $\delta \to 0$ and using the assumption that $\varrho(R) \leq 1-\epsilon$, we obtain $\lambda_1(J) \leq 2/\epsilon$.

Next, consider the conditional measure μ^τ with a pinning τ on a subset $U \subset V$. Let R^τ be the Dobrushin dependency matrix for μ^τ; note that by definition $R^\tau(x,y) = 0$ if $x \in U$ or $y \in U$, and $R^\tau(x,y) \leq R(x,y)$ for all $x, y \in V$. We deduce from Lemma 5.3.7 that $\varrho(R^\tau) \leq \varrho(R) \leq 1-\epsilon$ and thus this is reduced to the no-pinning case. Therefore, we get $\lambda_1(J^\tau) \leq 2/\epsilon$ for all τ and spectral independence then follows. □

5.3.3 Contraction for general Markov chains and general metrics

In this section, we generalize Theorem 5.3.2 to arbitrary "local" Markov chains and arbitrary metrics close to the Hamming metric. In particular, we prove Theorem 1.4.4.

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 5.3.8. 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 5.3.9. *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 5.3.2 and 1.4.3(2). For an arbitrary pinning τ and $(x, a) \in \mathcal{X}^\tau$, we define

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

and

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

where $t^\tau(x,a;y,a') = \mathrm{sgn}(J^\tau(x,a;y,a'))$; these definitions are analogous to (5.3.1) and (5.3.2) with pinning τ. Let $\tau' = \tau \cup (x,a)$. Then we deduce from Lemma 5.3.4 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 1.4.3(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 5.3.10. 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 5.3.3 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

$$D = \max_{p_t} \max_{x} \sum_{B \in \mathcal{B}: x \in B} p_t(B), \tag{5.3.8}$$

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 5.3.9 we establish Theorem 1.4.4, which we restate here for convenience.

Theorem 1.4.4. *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 5.3.9. 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 5.3.11. 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 1.4.4 the maximum block size M can be replaced by the maximum component size of all blocks.

Application: flip dynamics for colorings

In this section we establish spectral independence for colorings utilizing Theorem 1.4.4.

Theorem 5.3.12. *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 1.4.4, we need a contractive Markov chain for sampling colorings of a graph. Vigoda considered the *flip dynamics* [129] and showed that it is contractive for the Hamming metric when the number of colors $q > \frac{11}{6}\Delta$. Recently, [38] 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 [129] and the recent improvement [38], 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 [38], where the authors established contraction of the flip dynamics using the path coupling method.

Lemma 5.3.13 ([38])**.** *Under the assumptions of Theorem 5.3.12, 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 [38] generalize naturally to list colorings. Also, 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 5.3.3 addressing the same issue for the Glauber dynamics and also Remark 5.3.10 for general select-update dynamics.

We give below the proof of Theorem 5.3.12.

Proof of Theorem 5.3.12. Observe that the flip dynamics belongs to the class of select-update dynamics, where the associated $\mathcal{B}$ is the collection of connected subsets of vertices. Since the flip parameters satisfy $p_s > 0$ only for $s \leq 6$, we have $M \leq 6$. Moreover, we have $D \leq \Delta^6/n$ since a vertex x is in the selected bicolored component $L_{\sigma^t}(y, a)$ only if $\mathrm{dist}(x, y) \leq 5$, which happens with probability at most Δ^6/n. The theorem then follows from Lemma 5.3.13 and Theorem 1.4.4. □

We conclude here with the proof of Theorem 1.4.8.

Proof of Theorem 1.4.8. By Theorem 5.3.12 the uniform distribution μ of proper colorings is spectrally independent. Then the results follows immediately from Theorem 1.4.1. □

Application: block dynamics for Potts model

Here we apply Theorems 5.3.2 and 1.4.4 to the ferromagnetic Potts model to establish spectral independence.

Theorem 5.3.14. *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 [128] and [19] regarding the contraction of the Glauber dynamics and of the heat-bath block dynamics with a specific choice of blocks.

Lemma 5.3.15 ([128, Corollary 2.14] & [19, Proposition 2.2])**.** *Under the assumptions in Part 1 of Theorem 5.3.14, 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 5.3.16 ([19, Theorem 2.7])**.** *Under the assumptions in Part 2 of Theorem 5.3.14, 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 5.3.17. To be more precise, [19] shows that the conclusion of Lemma 5.3.16 is true when β, q, and the maximum block size $M = \max_{x \in V} |B_x|$ satisfy

$$\beta\left(\Delta - 1 + \frac{1}{M}\right) + 3M(\ln\Delta + \ln M) \leq \ln q. \tag{5.3.9}$$

Thus, for any $\delta > 0$, by taking $M = \lceil \delta^{-1} \rceil$ and $c = 3M(\ln\Delta + \ln M)$, our assumption $\beta \leq \frac{\ln q - c}{\Delta - 1 + \delta}$ in Part 2 of Theorem 5.3.14 implies (5.3.9). 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 $\beta \leq (1 - o(1))\frac{\ln q}{\Delta - 1}$ where $o(1)$ tends to 0 as $q \to \infty$; this gives the bound β_1 in Theorem 1.4.9 from the introduction.

Theorem 5.3.14 is an immediate consequence of Lemmas 5.3.15, 5.3.16 and the results proved in this section.

Proof of Theorem 5.3.14. Part 1 follows from Lemma 5.3.15 and Theorem 5.3.2. For Part 2, we note that the block dynamics from Lemma 5.3.16 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 5.3.16 and Theorem 1.4.4. □

We end this section with the proof of Theorem 1.4.9.

Proof of Theorem 1.4.9. For Ising model, spectral independence is known in the whole uniqueness region [41]. For Potts model, Theorem 5.3.14 establishes spectral independence in the corresponding parameter regimes. The theorem then follows from Theorems 1.4.1 and 1.4.2. □

5.4 Spectral independence and factorization of the entropy

The goal of this section is to reformulate in the setting of spin systems some of the key facts that were derived in [42] 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 [61, 25, 22], 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 [61, 25, 22], 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 [42].

Theorem 5.4.1. *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 $f \geq 0$:*

$$\frac{n}{\ell} \mathrm{Av}_{|U|=\ell} \mathrm{Ent}(\mu_{V\setminus U} f) \leq C \, \mathrm{Ent} f. \tag{5.4.1}$$

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

$$\frac{\ell}{n} \mathrm{Ent} f \leq C \, \mathrm{Av}_{|\Lambda|=\ell} \, \mu \left[\mathrm{Ent}_\Lambda f \right]. \tag{5.4.2}$$

We remark that, when $\ell = 1$, (5.4.1) takes the form of an approximate subadditivity statement:

$$\sum_{x \in V} \mathrm{Ent}(f_x) \leq C \, \mathrm{Ent}(f), \tag{5.4.3}$$

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 (5.4.3) is known to be equivalent to a Brascamp-Lieb type inequality for the measure μ; see [33, 32]. For a general discussion of subadditivity of entropy, Brascamp-Lieb type inequalities, and their applications, see for instance [7] and the references therein. On the other hand (5.4.2) is the uniform block factorization statement ℓ-UBF with $\ell = \lceil \theta n \rceil$; see Definition 5.1.1.

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 5.4.3. The second step is a control of the local inequality; see Lemma 5.4.4.

5.4.1 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 $\mu^\tau \left[\mathrm{Ent}_{\mu^{\tau,x,y}} f \right]$ indicates the entropy of f with respect to $\mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$,

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

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

$$(1 + \alpha_k) \mathrm{Av}_{x \notin U} \, \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \leq \mathrm{Av}_{x,y \notin U} \, \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \,, \tag{5.4.4}$$

hold for all $k = 0, \ldots, n-2$, for all $U \subset [n]$ with $|U| = k$, for all configurations τ on U and for all functions $f \geq 0$. The symbol $\mathrm{Av}_{x \notin U}$ denotes the uniform average over all $n - k$ vertices $x \notin U$, and $\mathrm{Av}_{x,y \notin U}$ 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 (5.4.4) as the local inequality, since for each choice of x, y, the distributions involved are concerned with the spins at two vertices only.

Remark 5.4.2. Fix $x, y \notin U$. Using $\mu^{\tau,x} f = \mu^{\tau,x} \mu^{\tau,x,y} f$, from Lemma 2.4.1 we have

the decomposition

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

In particular, $\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \geq \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))$ and therefore (5.4.4) 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 (5.4.4) with $\alpha_k = 1$ for all $k = 0, \dots, n-2$.

The recursion is based on the following statement, which rephrases [42, Theorem 5.4].

Lemma 5.4.3. *Let α_k, $k = 0, \dots, n-2$, be defined by* (5.4.4)*. Then, for all functions $f \geq 0$,*

$$\mathrm{Av}_{|U|=j}\mathrm{Ent}(\mu_{V\setminus U}f) \leq (1-\kappa_j)\mathrm{Ent}(f), \qquad j = 1, \dots, n-1, \tag{5.4.5}$$

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 (5.4.5) follows from the fact that for all $k = 1, \dots, n-1$:

$$\mathrm{Av}_{|U|=k}\,\mathrm{Ent}(\mu_{V\setminus U}f) \leq \delta_k \mathrm{Av}_{|U|=k+1}\,\mathrm{Ent}(\mu_{V\setminus U}f), \qquad \delta_k = \frac{\sum_{i=0}^{k-1}\Gamma_i}{\sum_{i=0}^{k}\Gamma_i}, \tag{5.4.6}$$

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 (5.4.6), note that it holds for $k=1$ with $\delta_1 = 1/(1+\alpha_0) = \Gamma_0/(\Gamma_0+\Gamma_1)$ by the assumption (5.4.4) 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 2.4.1 we have the decomposition

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

Averaging we obtain

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

In the same way

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

From (5.4.4),

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

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 (5.4.6) at $k-1$ we have

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

□

5.4.2 Estimating the local coefficients

The next step is an estimate on the coefficients α_k appearing in (5.4.4).

Lemma 5.4.4. *If the spin system is η-spectrally independent and b-marginally bounded then the local inequality* (5.4.4) *holds with*

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

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

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

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

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \\
&\quad \geq 1 - \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x}(f)\mu^{\tau,y}(f)\right] \\
&\quad = -\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right]. \qquad (5.4.7)
\end{aligned}$$

We may rewrite

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right] \\
&\qquad = \frac{1}{n-k-1}\sum_{(x,a)\in\mathcal{X}} \nu(x,a)\varphi(x,a)[J^\tau\varphi](x,a), \qquad (5.4.8)
\end{aligned}$$

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 2.2.5. Note that in the derivation of (5.4.8) 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 $L^2(\mathcal{X},\nu)$:

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

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,

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

Recalling (5.4.7) we have shown

$$\begin{aligned}\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \\ \geq -\frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).\end{aligned} \tag{5.4.9}$$

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 2.2.7, 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

$$\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)) \leq \frac{2}{b}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)). \tag{5.4.10}$$

Inserting (5.4.10) into (5.4.9) concludes the proof. □

5.4.3 Proof of Theorem 5.4.1

From Lemma 5.4.3, we see that (5.4.1) holds with $C = \frac{n}{\ell}(1-\kappa_\ell)$. From Lemma 5.4.4 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 [42], it is not hard to see that for any $1 \leq \ell \leq n-1$:

$$\kappa_\ell \geq \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}. \tag{5.4.11}$$

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 (5.4.1) holds with $C \le R+1 = O(1+\frac{\eta}{b})$.

To prove (5.4.2), 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 2.4.1. Therefore Lemma 5.4.3 implies that (5.4.2) holds with $C = \frac{\ell}{n\,\kappa_{n-\ell}}$. Using (5.4.11) we see that

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

In particular, if $\ell = \lceil \theta n \rceil$ with $\theta \in (0,1]$ fixed, then for all sufficiently large n one has $\frac{\ell}{n\,\kappa_{n-\ell}} \le (\frac{1}{\theta})^{O(R)}$. This ends the proof of Theorem 5.4.1.

5.5 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 (5.1.2), 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 1.4.2 from the introduction.

We again take $G = (V, E)$ to be an n-vertex graph of maximum degree Δ, μ to be the Potts distribution on G with configuration space $\Omega = [q]^V$ and ν to be the Edward-Sokal measure. We refer to section 4.1 for the definition of ν and the spin/edge factorization.

The main result in this section is stated as follows.

Theorem 5.5.1. *Suppose μ satisfies the k-partite factorization of entropy with constant C_{par}; see Eq.* (5.1.2)*. Then, there exists a constant $C = C(C_{\mathrm{par}}, \beta, \Delta)$ such that for all $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$*

$$\mathrm{Ent}_\nu(f) \le C\ \left(\nu\left[\mathrm{Ent}_\nu(f \mid \sigma)\right] + \nu[\mathrm{Ent}_\nu(f \mid A)]\right).$$

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

Theorem 1.4.2 from the introduction now follows immediately.

Proof of Theorem 1.4.2. For the Potts model one has $e^{\beta\Delta} = O(1/b)$. Therefore, the results follows from Theorem 5.4.1, Lemma 5.1.3, Theorem 5.5.1 and Lemma 4.1.4. □

Let $\{V_1, ..., V_k\}$ be the k-partition of G, where $k \leq \Delta + 1$, as in Section 5.2. 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 $\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)$ for the corresponding conditional entropy and $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A)\right]$ for its expectation with respect to ν. Theorem 5.5.1 will follow from the following lemmas.

Lemma 5.5.2. *For all $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$ and all $j \in [k]$ we have*

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

Lemma 5.5.3. *There exists a constant $\delta_1 > 0$ such that, for all $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$ and all $j \in [k]$,*

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

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

Lemma 5.5.4. *If μ satisfies the k-partite factorization with constant C_{par}, then for all $f : \Omega_{\mathrm{J}} \mapsto \mathbb{R}_+$,*

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

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

Proof of Theorem 5.5.1. By combining the bounds from Lemmas 5.5.2, 5.5.3 and 5.5.4 we get

$$\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma) + \mathrm{Ent}_\nu(f \,|\, A)\right] \geq \frac{\delta_1\delta_2}{k} \mathrm{Ent}_\nu(f),$$

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

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

We turn to the proof of Lemmas 5.5.2, 5.5.3 and 5.5.4. In the special case of bipartite graphs these correspond to Lemmas 4.3.3, 4.3.4 and 4.3.5 in chapter 4, respectively. For Lemmas 5.5.2 and 5.5.4 the adaptation to our setting is straightforward. The proof of Lemma 5.5.3, the core of the argument, requires some modification. The main difference with the proof of lemma 4.3.4 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 5.5.2. This is an instance of the same monotonicity already seen in Lemma 2.4.2. In this particular case, it follows from the argument in the proof of 4.3.3 by simply substituting σ_O with $\sigma_{V_j^c}$ in that proof. □

Proof of Lemma 5.5.3. 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

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

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:

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

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

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

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 $x \in V$,

$$\nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma) \,|\, \sigma_{V_j^c} \right] + \nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c} \right] \geq \delta_1 \, \mathrm{Ent}_x(f \,|\, \sigma_{V_j^c}). \quad (5.5.1)$$

In the case $x \in V_j$ this follows exactly as in Lemma 4.3.6 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 want to lift inequality (5.5.1) 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

$$\mathrm{Ent}_\nu(f \,|\, \sigma_{V_j^c}) = \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}) \,|\, \sigma_{V_j^c} \right], \quad (5.5.2)$$

where $g_x = \nu\left[f \,|\, \sigma_{V_j^c}, \xi_{x+1}, \ldots, \xi_n\right]$, $g_0 = f$ and $g_n = \nu\left[f \,|\, \sigma_{V_j^c}\right]$. This identity is an instance of the decomposition in Lemma 2.4.1.

Putting together (5.5.1) and (5.5.2) yields

$$\begin{aligned}\delta_1 \operatorname{Ent}_\nu(f \,|\, \sigma_{V_j^c}) &\le \sum_{x=1}^{n} \nu\left[\nu_x\left[\operatorname{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \sigma_{V_j^c}\right] + \nu_x\left[\operatorname{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c}\right] \,|\, \sigma_{V_j^c}\right] \\ &= \sum_{x=1}^{n} \nu\left[\operatorname{Ent}_x(g_{x-1} \,|\, \sigma) + \operatorname{Ent}_x(g_{x-1} \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c}\right]. \qquad (5.5.3)\end{aligned}$$

To conclude the proof we can now proceed exactly as in the proof of Lemma 4.3.4. We obtain the following two inequalities:

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

These two inequalities combined with (5.5.3) yield that

$$\delta_1 \operatorname{Ent}_\nu(f \,|\, \sigma_{V_j^c}) \le \nu\left[\operatorname{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_{V_j^c}\right] + \nu\left[\operatorname{Ent}_\nu(f \,|\, \sigma_{V_j^c}, A) \,|\, \sigma_{V_j^c}\right]. \qquad (5.5.4)$$

The desired result follows by taking expectations with respect to ν in (5.5.4). □

Proof of lemma 5.5.4. From the definition of conditional entropy and the fact that $\nu(\cdot \,|\, \sigma_{V_j}, \sigma_{V_j^c}) = \nu(\cdot \,|\, \sigma)$ we get

$$\operatorname{Ent}_\nu(f \,|\, \sigma_{V_j^c}) = \operatorname{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right] \,|\, \sigma_{V_j^c}\right) + \nu\left[\operatorname{Ent}_\nu(f \,|\, \sigma) \,|\, \sigma_{V_j^c}\right]. \qquad (5.5.5)$$

(see eq. (4.3.4),(4.3.5) from Lemma 4.3.5). Now, since the function $\nu\left[f \,|\, \sigma\right]$ depends only on the spin configuration σ, one has the identity

$$\sum_{j=1}^{k} \nu\left[\operatorname{Ent}_\nu(\nu[f \,|\, \sigma] \,|\, \sigma_{V_j^c})\right] = \sum_{j=1}^{k} \mu\left[\operatorname{Ent}(\nu[f \,|\, \sigma] \,|\, \sigma_{V_j^c})\right], \qquad (5.5.6)$$

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,

$$\sum_{j=1}^{k} \mu\left[\operatorname{Ent}(\nu[f \,|\, \sigma] \,|\, \sigma_{V_j^c})\right] \ge \delta_2 \operatorname{Ent}\left(\nu\left[f \,|\, \sigma\right]\right), \qquad (5.5.7)$$

where $\delta_2 = 1/C_{\text{par}}$. 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 (5.5.5), and combining with (5.5.6) and (5.5.7), 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

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

5.6 Open problem: block factorization for mean field spin system

In this section we briefly discuss an open problem regarding the block factorization (1.2.4). So far, we discussed its validity for spin systems where the maximum degree of G is bounded. An interesting question is whether it is possible to remove the assumption of bounded degree, that is prove block factorization with a constant C independent on $|V|$ for Gibbs measures on arbitrary graphs assuming spectral independence and marginal boundedness only. A well known example of Gibbs measure on a graph whose maximum degree is n is the Curie-Weiss model. Let $G = (V, E)$ be the $n-$complete graph, and $\Omega = \{-1, 1\}^n$; the ferromagnetic Curie-Weiss model is described by the following distribution

$$\mu(\sigma) = \exp\left(\frac{\beta}{n} \sum_{\{i,j\} \in E} \sigma_i \sigma_j - h \sum_{i \in V} \sigma_i \right), \qquad \sigma \in \Omega,$$

where $\beta \geq 0$ denote the inverse temperature of the system and $h \geq 0$ denotes the external magnetic field. Here the uniqueness region is $\beta < 1$. There are results in the literature concerning the approximate tensorization statement for this model. It holds, for example, assuming a version of Dobrushin's uniqueness condition, see [101]. More recently, in [3] and [39] have been introduced two powerful tools, entropic independence and stochastic localization respectively. These techniques are used to attack the approximate tensorization statement for the Curie Weiss and also other far more complex models, such as the Sherrington-Kirkpatrick model. However, as far as we know there isn't any statement concerning the validity of the block factorization for these cases, even for the simpler Curie-Weiss. In this latter case, the best result one has so far is $C = O(n)$ if the Dobrushin's uniqueness condition in [101] holds. Note that this condition holds if $\beta < 1$,

so that all the uniqueness regime is covered. Below we prove how to derive $C = O(n)$ for the Curie-Weiss.

Recall that from Lemma 2.4.2 we have

$$\mu\left[\mathrm{Ent}_x(f)\right] \leq \mu\left[\mathrm{Ent}_B(f)\right]$$

for any $B \subset V$ and $x \in B$. From this one can deduce that the following inequality

$$\frac{1}{|A|}\sum_{x\in A}\mu\left[\mathrm{Ent}_x(f)\right] \leq \mu\left[\mathrm{Ent}_A(f)\right]$$

holds for any $A \subset [n]$. By using approximate tensorization we get

$$\begin{aligned}\frac{\gamma(\alpha')}{C_{AT}}\mathrm{Ent}_V(f) &\leq \sum_{x\in[n]}\sum_{A\subset V}\alpha_A\frac{1}{|A|}\mathbf{1}(x\in A)\mu\left[\mathrm{Ent}_x(f)\right]\\ &\leq \sum_{A\subset V}\alpha_A\mu\left[\mathrm{Ent}_A(f)\right],\end{aligned}$$

so that the best constant C such that (1.2.4) holds is upper bounded by

$$C \leq \frac{\gamma(\alpha)}{\gamma(\alpha')}C_{AT} \leq nC_{AT}$$

where $\gamma(\alpha) := \inf_{x\in[n]}\sum_{\substack{A\subset V\\ x\in A}}\alpha_A$, $\gamma(\alpha') = \inf_{x\in[n]}\sum_{\substack{A\subset V\\ x\in A}}\frac{\alpha_A}{|A|}$ and C_{AT} is the best constant such that approximate tensorization holds, which we know is $O(1)$ thanks to the main result in [101].

Part II

Non-linear recombination models

Chapter 6

Introduction

Recombinations are one of the principal components in the analysis of stochastic genetic algorithms [92, 109]. Nonlinear recombinations provide a simple combinatorial setup for quadratic evolutions described by a Boltzmann-like equation [115]. A particle is represented by a finite string of characters from some finite alphabet and the binary collision mechanism is given by a recombination, that is the transposition of a random portion of the two colliding strings. The model belongs to the family of symmetric quadratic systems introduced in [116]; see also [29] for the more general framework of reversible quadratic systems.

Following the strategy introduced by Mark Kac in his seminal 1956 paper [82], one can approximate the nonlinear evolution of one particle by a linear mean field type Markov process involving a large number of particles. Roughly speaking, if one has a good control of this approximation, together with a good control of the linear particle system, then the difficulties due to the nonlinearity in the original process can be overcome.

In the context of Boltzmann's equation and its closely related kinetic models this line of research has witnessed important progress in recent years [31, 106, 107, 37], see also [34, 52, 83] for related results for mean field diffusions of McKean-Vlasov type.

The combinatorial setup considered here appears to be less explored; see however [113, 118, 8] for the analysis of Boltzmann-like equations with discrete velocities. One advantage of the combinatorial setting is that thanks to the discrete setup one can avoid a number of technical assumptions, such as regularity and moments constraints, on the various distributions considered. Moreover, and perhaps more importantly, in contrast with the well studied case of the Kac-Boltzmann equation [31, 106, 57, 18], in our setup it is possible to obtain tight entropy production estimates for the particle system which hold uniformly in the number of particles. This provides a class of models for which the renowned Kac

program can be completed in a strong sense.

In the setting of nonlinear recombinations the linear particle system takes the form of a generalized random transposition dynamics. This yields a natural generalization of the mean field exchange dynamics that are commonly studied in the probabilistic literature such as the Bernoulli-Laplace or the random transposition model [48]. The purpose is twofold. On one hand we establish *uniform in time propagation of chaos*. On the other hand we prove tight estimates on the *entropy production* of the linear system which hold *uniformly in the number of particles*. As a corollary we obtain quantitative control on the convergence to stationarity for the nonlinear model in terms of relative entropy. In particular, this extends some results previously obtained in [115, 29] by direct analysis of the entropy in the nonlinear recombination model. We now proceed with a detailed description of the model and of our main results.

6.1 The nonlinear equation

Let $\Omega = \prod_{i=1}^{n} X_i$ be the set of n-vectors $\sigma = (\sigma_1, \ldots, \sigma_n)$ where $\sigma_i \in X_i$, and the X_i are given finite sets. We interpret σ as a particle. Thus, a particle is a string of n characters each taken from a finite alphabet. A basic example is obtained by taking $\Omega = \{0,1\}^n$. Without loss of generality we will assume that each X_i has the form $X_i := \{0, 1, 2, \ldots, q_i\}$, for some $q_i \in \mathbb{N}$. Given a subset $A \subset [n]$, $[n] = \{1, \ldots, n\}$, and $\sigma \in \Omega$, σ_A denotes the $A-$component of σ, that is the string $(\sigma_i, i \in A)$. If $(\sigma, \eta) \in \Omega \times \Omega$ is a pair of particles, the recombination at A consists in exchanging the A-component of σ with the A-component of η. This defines the map

$$(\sigma, \eta) \mapsto (\eta_A \sigma_{A^c}, \sigma_A \eta_{A^c}),$$

where $\eta_A \sigma_{A^c}$ denotes the element of Ω with entries η_i for $i \in A$ and σ_i for $i \in A^c = [n] \setminus A$. Let $\mathcal{P}(\Omega)$ denote the set of probability measures on Ω. If the original pair (σ, η) is obtained by sampling independently from $p \in \mathcal{P}(\Omega)$, then the new particle $\eta_A \sigma_{A^c}$ is distributed according to $p_A \otimes p_{A^c}$, the product of the two marginals of p on the A and A^c components respectively. By choosing the set $A \subset [n]$ according to some distribution ν, one obtains the quadratic collision kernel

$$p \mapsto Q(p) = \sum_{A \subset [n]} \nu(A)\, p_A \otimes p_{A^c}.$$

The nonlinear evolution is defined by the dynamical system $\dot{p}_t = Q(p_t) - p_t$, that is

$$\frac{d}{dt}\, p_t = \sum_{A \subset [n]} \nu(A)\, (p_{t,A} \otimes p_{t,A^c} - p_t) \tag{6.1.1}$$

with the initial condition $p_0 = p \in \mathcal{P}(\Omega)$. Here $p_t \in \mathcal{P}(\Omega)$ is the distribution of the particle at time t, $p_{t,A}$ denotes its marginal on A and ν is a given probability measure over the subsets of $[n]$. The study of this model starts with the pioneering work of Geiringer [65]; see also [116, 115, 6, 100] for more recent accounts. It is well known that the Cauchy problem associated to (6.1.1) has a unique solution for every initial distribution $p \in \mathcal{P}(\Omega)$. Moreover, it is not difficult to see that the evolution preserves the single site marginals, that is $p_{t,i} = p_i$ for all $t \geq 0$ and for all $i \in [n]$. We say that the recombination measure ν is *separating* if for any $i, j \in [n]$ there is a positive probability that the random set A with distribution ν separates i and j, namely if $r(\nu) < 1$ where we define

$$r(\nu) := \max_{i,j \in [n]} \mathrm{Pr}_\nu \left(\{i, j\} \subset A \text{ or } \{i, j\} \subset A^c \right). \tag{6.1.2}$$

It is a classical fact that, under the assumption that ν is separating, the system converges to the stationary state given by the product of the marginals of the initial state p; namely, if $\pi_i = p_i$ denotes the marginal of p at site i, then

$$\pi = \otimes_{i=1}^n \pi_i \tag{6.1.3}$$

is the equilibrium distribution and one has the convergence $p_t \to \pi$, $t \to \infty$, which can be interpreted as the effect of repeated fragmentations of the initial state. Some of our results will hold for arbitrary separating ν. In some other cases we consider a slightly stronger assumption on ν. Two examples to which all our results apply are the following distributions ν, which are commonly considered in the genetic recombination literature:

1. *Uniform crossover*: $\nu(A) = \frac{1}{2^n}$, for all $A \subset [n]$;
2. *One-point crossover*: $\nu(A) = \frac{1}{n+1} \sum_{i=0}^n \mathbf{1}_{A = J_i}$, where $J_0 = \emptyset$, $J_i = \{1, \ldots, i\}$, $i \geq 1$.

The quantitavive analysis of the convergence to equilibrium $p_t \to \pi$, $t \to \infty$ has been initiated in [116, 115], where a "mixing time" bound was obtained for the discrete time version of the model. The decay to equilibrium in relative entropy for the continuous time model was studied in [29]. These results were obtained by direct analysis of the nonlinear problem. Here we shall follow an entirely dif-

ferent approach, inspired by Kac's program from kinetic theory. As a byproduct of our analysis, we shall obtain an alternative proof of the known results mentioned above.

6.2 The particle system

Suppose there are N "particles", described by variables $\eta(j) \in \Omega$, $j = 1, \dots, N$. That is, each particle is a single string from Ω and $\eta_i(j)$ denotes the content of the j-th particle at site $i \in [n]$. We may picture $\eta \in \Omega^N$ as a $N \times n$ matrix such that each row is a particle with n entries, and for each $i \in [n]$, the i-th column η_i represents the content of site i for different particles.

Notice that N and n play two very different roles here. The number N of particles will eventually be taken to $+\infty$ to recover the non linear mean field limit, in accordance with the general Kac program. The number n should be thought as a fixed, possibly large quantity describing the size of a single particle space.

The Markov process is given by the following random pair-exchange process. Pairs of particles $\{j, \ell\}$, $1 \leq j < l \leq N$ are chosen independently according to a Poisson clock process with rate $1/N$. When the pair $\{j, l\}$ "rings", then a set $A \subset [n]$ is chosen with probability $\nu(A)$ and the recombination

$$(\eta(l), \eta(j)) \mapsto (\eta_A(j)\eta_{A^c}(l), \eta_A(l)\eta_{A^c}(j)) \tag{6.2.1}$$

is performed, that is the A-content is exchanged between particle j and particle l. For all $j, l \in [N]$, $A \subset [n]$, for all $\eta \in \Omega^N$, we write $\eta^{j,l,A}$ for the new configuration $\eta' \in \Omega$ defined by

$$\begin{aligned} \eta'(k) &= \eta(k), \ \forall k \neq j, l; \\ \eta'(l) &= \eta_A(j)\eta_{A^c}(l), \\ \eta'(j) &= \eta_A(l)\eta_{A^c}(j). \end{aligned} \tag{6.2.2}$$

With this notation (6.2.1) can be rewritten as

$$(\eta(l), \eta(j)) \mapsto (\eta^{j,l,A}(l), \eta^{j,l,A}(j)).$$

Define also $f^{j,l,A}(\eta) = f(\eta^{j,l,A})$ for all $f : \Omega^N \to \mathbb{R}$. Then, the infinitesimal generator of the continuous time Markov process is given by

$$\mathcal{L}_N f = \frac{1}{N} \sum_{1 \leq j < l \leq N} \sum_{A \subset [n]} \nu(A) \left(f^{j,l,A} - f \right), \qquad f : \Omega^N \to \mathbb{R}.$$

Any product measure μ_N on Ω^N of the form $\mu_N = \mu^{\otimes N}$, where μ is itself a product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ on $\Omega = \prod_{i=1}^n X_i$, defines a reversible measure for the generator $\mathcal{L}_N$. Indeed, for such a measure one has the symmetry $\mu_N(\eta^{j,\ell,A}) = \mu_N(\eta)$ for all $\eta \in \Omega^N$ and therefore $\mathcal{L}_N$ is self-adjoint in $L^2(\mu_N)$. The process is not irreducible in the state space Ω^N since the content at a site for one particle is always exchanged with the content at the same site for another particle, and thus the number of particles with a given element $x \in X_i$ at a given site $i \in [n]$ is constant in time. To obtain an irreducible process one must fix the densities $\varrho_{i,x}$, $i \in [n], x \in X_i$ defined by

$$\varrho_{i,x} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(\eta_i(j) = x).$$

We call $\varrho = (\varrho_{i,x})$ the corresponding vector. Given a density vector ϱ, consider the space

$$\Omega_\varrho := \left\{ (\eta(1), \dots, \eta(N)) \in \Omega^N \;\middle|\; \begin{array}{l} \varrho_{i,x} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(\eta_i(j) = x), \\ \forall i \in [n], x \in X_i \end{array} \right\}. \tag{6.2.3}$$

The set Ω_ϱ is well defined and non-empty for every vector $\varrho = \varrho_N$ such that $\varrho_{i,x} \in [0,1]$, $\sum_{x \in X_i} \varrho_{i,x} = 1$ for all $i \in [n]$, and such that $N\varrho_{i,x}$ is an integer for all i, x. When this holds we say that ϱ_N is an *admissible sequence*. Under suitable assumptions on the recombination measure ν, see Definition 6.5.1, for any given admissible ϱ_N, the Markov process with state space Ω_{ϱ_N} and generator $\mathcal{L}_N$ is irreducible and converges to the uniform distribution on Ω_{ϱ_N}. We will be interested in quantitative statements about this convergence.

We often use the following procedure to construct admissible sequences. Fix a given $\pi = (\pi_{i,x})$ satisfying $\pi_{i,x} \in [0,1]$ for all i, x and $\sum_{x \in X_i} \pi_{i,x} = 1$ for all $i \in [n]$. Then we call $\varrho^\pi = \varrho^\pi(\pi)$ the density defined by

$$\varrho^\pi_{i,x} := \tfrac{1}{N} \lfloor N\pi_{i,x} \rfloor, \qquad i \in [n],\ x \in \{1, \dots, q_i\}, \tag{6.2.4}$$

and we set $\varrho^\pi_{i,0} = 1 - \sum_{x=1}^{q_i} \varrho^\pi_{i,x}$. We remark that $\varrho^\pi = \varrho^\pi_N$ is an admissible sequence satisfying

$$\varrho^\pi_{i,x} = \pi_{i,x} + O\left(\tfrac{1}{N}\right)$$

for all i, x. Fixing the probability vector $\pi = (\pi_{i,x})$ is equivalent to fixing the stationary measure (6.1.3) of the nonlinear evolution, and thus we use the same symbol for them. From now on it is assumed that the densities π, and thus the corresponding product measure π, are fixed. Without loss of generality we restrict to the case where $\pi_{i,x} \in (0,1)$ for all $i \in [n], x \in X_i$ since for each i we can

otherwise discard those letters $x \in X_i$ such that $\pi_{i,x}$ is zero, and thus consider a new configuration space such that π is everywhere positive.

6.3 Chaos and the propagation of chaos

The chaos property is commonly defined as follows. A measure $\mu_N \in \mathcal{P}\left(\Omega^N\right)$ is *symmetric* if it is invariant under any permutation of the N particles. We write $P_k\mu_N \in \mathcal{P}(\Omega^k)$ for the corresponding k-particle marginal.

Definition 6.3.1 (Kac's chaos or "Boltzmann property")**.** A sequence $\mu_N \in \mathcal{P}\left(\Omega^N\right)$ of symmetric probabilities on Ω^N is $\mu-$chaotic, for a given $\mu \in \mathcal{P}(\Omega)$, if for any $k \in \mathbb{N}$ one has the weak convergence

$$P_k\mu_N \longrightarrow \mu^{\otimes k}\,, \qquad N \to \infty.$$

A key step in implementing Kac's program is to construct a correspondence between probability measures on Ω and probability measures on Ω^N. In our setting this can be formulated as follows.

Definition 6.3.2 (Canonical tensor product)**.** Given a probability measure $p \in \mathcal{P}(\Omega)$ and an admissible sequence of density vectors ϱ_N, we let

$$\gamma(p, \varrho_N) := p^{\otimes N}\left(\,\cdot\,|\,\Omega_{\varrho_N}\right)$$

be the tensor product of p conditioned on Ω_{ϱ_N}. When ϱ_N is given by ϱ^π as in (6.2.4), where the $\pi_{i,x} = p(\sigma_i = x)$ are the marginals of p, we use the notation $\gamma_N(p) := \gamma(p, \varrho^\pi)$, and call it the *canonical tensor product*.

To avoid degeneracies we sometimes assume the following property.

Definition 6.3.3 (Irreducibility)**.** A probability measure $p \in \mathcal{P}(\Omega)$ is called *irreducible* if for any $i \in [n]$, any $x \in \{1, \dots, q_i\}$, there exists $\chi \in \Omega$ such that $p(\sigma_i = x, \sigma_j = \chi_j\ \forall j \neq i) > 0$ and $p(\sigma_i = 0, \sigma_j = \chi_j\ \forall j \neq i) > 0$.

If p is irreducible and the sequence ϱ_N is sufficiently close to the marginals of p, then the local central limit theorem guarantees that the k-particle marginals of the symmetric measures $\gamma(p, \varrho_N)$ converge to the product $p^{\otimes k}$ as $N \to \infty$, that is $\gamma(p, \varrho_N)$ is p-chaotic, see Theorem 7.1.6 for a precise statement. In particular, for the canonical tensor product $\gamma_N(p)$, we will see that

$$\|P_k\gamma_N(p) - p^{\otimes k}\|_{\text{TV}} \leq \frac{C_0\, k}{N}, \tag{6.3.1}$$

for some constant $C_0 = C_0(p)$. Clearly, our reference product measure $\pi \in \mathcal{P}(\Omega)$ is irreducible. In fact, $\gamma(\pi, \varrho_N)$ is the *uniform probability measure on* Ω_{ϱ_N}, for any admissible sequence ϱ_N. In particular, it follows that its k-particle marginals converge to $\pi^{\otimes k}$ as $N \to \infty$. One can also show that $\gamma(p, \varrho_N)$ is *entropically p-chaotic* in the sense defined in [31], namely that on top of the convergence of marginals one also has

$$\lim_{N\to\infty} \frac{1}{N} H_N(\gamma(p, \varrho_N) \,|\, \gamma(\pi, \varrho_N)) = H(p \,|\, \pi),$$

where $H(\cdot\,|\,\cdot)$, $H_N(\cdot\,|\,\cdot)$ denote respectively the relative entropy for probability measures on Ω and on Ω^N, see Proposition 7.1.8. Let us recall the following standard definition.

Definition 6.3.4 (Propagation of chaos)**.** Let $\mu_{N,t} = \mu_N e^{t\mathcal{L}_N}$, $t \geq 0$, denote the evolution of an initial symmetric distribution $\mu_N \in \mathcal{P}(\Omega^N)$ under the Markov process generated by $\mathcal{L}_N$. Suppose that μ_N is p-chaotic for some $p \in \mathcal{P}(\Omega)$ and let p_t denote the evolution of the initial datum p under the nonlinear process (6.1.1). If $\mu_{N,t}$ is p_t-chaotic for all fixed $t \geq 0$, then we say that *propagation of chaos* holds. If the weak convergence $P_k\mu_{N,t} \longrightarrow p_t^{\otimes k}$, $N \to \infty$, holds uniformly in $t \geq 0$ we say that propagation of chaos holds *uniformly in time*.

An adaptation of well known arguments, see e.g. [82, 125], shows that the propagation of chaos (at fixed times) holds in our setting. In fact, the proof of this does not require the assumption that ν is separating.

6.4 Uniform in time propagation of chaos

Our first main result concerns the validity of propagation of chaos uniformly in time, with quantitative bounds on the convergence as $N \to \infty$.

Theorem 6.4.1. *Assume that ν is separating. The propagation of chaos holds uniformly in time, that is for any $p \in \mathcal{P}(\Omega)$, if μ_N is $p-$chaotic, then for all fixed $k \in \mathbb{N}$, as $N \to \infty$,*

$$P_k\mu_{N,t} \longrightarrow p_t^{\otimes k}, \qquad \text{uniformly in } t \geq 0,$$

where p_t is the solution to the nonlinear equation (6.1.1) *with initial datum $p_0 = p$. Moreover, if μ_N is the canonical tensor product $\mu_N = \gamma_N(p)$, and $p \in \mathcal{P}(\Omega)$ is irreducible, then*

$$\|P_k\mu_{N,t} - p_t^{\otimes k}\|_{\text{TV}} \leq \frac{C}{\sqrt{N}}, \tag{6.4.1}$$

for some constant $C = C(k, p) > 0$ independent of t, N.

Remark 6.4.2. Concerning the dependency on N it may be that the optimal decay in (6.4.1) is $O(1/N)$ rather than $O(1/\sqrt{N})$. This seems natural in light of our estimate (6.3.1) at time zero. Moreover, that would be in agreement with the recent results in [83], where the $O(1/N)$ bound is obtained for a class of interacting diffusion processes at fixed times. The value of the constant C in (6.4.1) can be in principle obtained from our more detailed results in Theorem 7.2.1. However, we have not tried to optimize the dependency of C on k, p.

There are by now several results for kinetic models and for mean field diffusions establishing uniform in time propagation of chaos, see [34, 52, 37, 120]. However, the adaptation to our setting of the different techniques used in these works does not seem to be straightforward. The proof of Theorem 6.4.1 is based on some new contractive estimates for the nonlinear model that allow us to implement the main strategy developed in the groundbreaking work of Mischler and Mouhot [106], see Section 7.2.

6.5 Entropy production for generalized random transpositions

Our second main result is about quantitative estimates on the decay to equilibrium for the particle system introduced above. The key feature is that these estimates hold *uniformly in the number of particles* N. We shall actually derive such estimates in the context of the *generalized random transposition process* defined as follows.

Let $\mathcal{S}_{N,n} = S_N^n$ denote the n-fold product of the symmetric group S_N of the permutations of $[N] = \{1, \dots, N\}$. Then $\eta \in \mathcal{S}_{N,n}$ is a matrix $\eta_i(j)$, $i \in [n]$, $j \in [N]$, where each $\eta(j) = (\eta_1(j), \dots, \eta_n(j)) \in [N]^n$ is seen as a particle, and each $\eta_i = (\eta_i(1), \dots, \eta_i(N)) \in S_N$ is a permutation of $[N]$. Note that $\mathcal{S}_{N,n} = \Omega_\varrho$ where Ω_ϱ is defined in (6.2.3) when we take the extreme case $q_i \equiv N - 1$, $\varrho_{i,x} \equiv 1/N$ for all $i = 1, \dots, n$.

The generalized random transposition (GRT) process is defined as the process generated by the operator $\mathcal{L}_N$ in this setup, namely the GRT process is the continuous time Markov process with state space $\mathcal{S}_{N,n}$ described as follows: every pair of particles $\{\eta(l), \eta(j)\}$ collides with rate $1/N$ independently, and when a collision occurs, a new set A is sampled according to ν and the A-content of $\eta(l), \eta(j)$ is exchanged.

This setting is convenient for proving functional inequalities since by restricting to classes of functions with suitable symmetries we then recover all possible cases of processes on Ω_{ϱ_N} with generator $\mathcal{L}_N$, for all admissible ϱ_N. As an exam-

ple, consider the case $q_i \equiv 1$ and suppose that $N(\varrho_N)_{i,1} = N_i$ for some positive integers N_i, $i = 1, \dots, n$. Here the process generated by $\mathcal{L}_N$ can be seen as the GRT process restricted to functions $f : \mathcal{S}_{N,n} \mapsto \mathbb{R}$ such that, for each i, f only depends on $(\eta_i(1), \dots, \eta_i(N))$ through the unordered set $\{\eta_i(1), \dots, \eta_i(N_i)\}$. This can be seen as a generalized Bernoulli-Laplace process [47]. If no confusion arises we continue to write $\mathcal{L}_N$ for the generator of the GRT.

We remark that when $n = 1$, GRT is just the usual random transposition process [48], and that when ν gives positive weight only to $A \subset [n]$ such that $|A| = 1$, it describes n independent random transposition processes. However, in the general case, the recombination measure ν dynamically couples the permutations and the GRT becomes a nontrivial generalization of the standard random transpositions.

In order to guarantee the irreducibility of the GRT process, we make the following assumption on the recombination measure ν, which is easily seen to be stronger than the separation assumption $r(\nu) < 1$; see also Remark 6.5.4.

Definition 6.5.1. We say that ν is *strictly separating* if for all $i \in [n]$ there exists $A \subset [n]$ such that $i \in A$ and such that both A and $A \setminus \{i\}$ have positive ν-probability.

Note that the uniform crossover and the one-point crossover are both strictly separating. Let π_N denote the uniform distribution on $\mathcal{S}_{N,n}$. The GRT process is reversible with respect to π_N and if the measure ν is strictly separating, then it is also irreducible, and any initial distribution converges to π_N as $t \to \infty$. To quantify this statement we consider the Dirichlet form of the GRT, defined by

$$\mathcal{E}_{N,n}(f,g) = \frac{1}{2N} \sum_{1 \le j < l \le N} \sum_{A \subset [n]} \nu(A) \sum_{\eta \in \mathcal{S}_{N,n}} \pi_N(\eta) \left(f(\eta^{j,l,A}) - f(\eta) \right) \left(g(\eta^{j,l,A}) - g(\eta) \right),$$

where $f, g : \mathcal{S}_{N,n} \mapsto \mathbb{R}$. The *entropy production rate* is measured by the constant

$$\alpha(N,n) = \inf_{f>0} \frac{\mathcal{E}_{N,n}(f, \log f)}{\mathrm{Ent}(f)}, \tag{6.5.1}$$

where the infimum is over $f : \mathcal{S}_{N,n} \mapsto \mathbb{R}_+$ such that $\mathrm{Ent}(f) \neq 0$ and

$$\mathrm{Ent}(f) = \pi_N(f \log f) - \pi_N(f) \log \pi_N(f)$$

is the entropy of f w.r.t. π_N. Equivalently, $\alpha(N,n)$ is the best constant α such that the inequality

$$\mathrm{Ent}(e^{t\mathcal{L}_N} f) \le e^{-\alpha t} \mathrm{Ent}(f)$$

holds for all functions $f > 0$; see e.g. [46, 17]. Note that when $\pi_N(f) = 1$ then, for all $t \geq 0$, $\mathrm{Ent}(e^{t\mathcal{L}_N} f)$ coincides with the relative entropy $H_N(\mu_{N,t} \,|\, \pi_N)$ where $\mu_{N,t} = (e^{t\mathcal{L}_N} f)\pi_N$.

We also consider the entropy production rate restricted to the set of symmetric functions defined as follows. Let $\mathbb{S}$ denote the set of $f : \Omega^N \mapsto \mathbb{R}$ such that

$$f(\eta) = \frac{1}{N!} \sum_{\tau \in S_N} f(\tau \circ \eta),$$

where the sum runs over all permutations $\tau \in S_N$ and $\tau \circ \eta$ denotes the configuration with particles exchanged according to τ, that is $\tau \circ \eta(j) = \eta(\tau(j))$. From the point of view of Kac's program [82], $\mathbb{S}$ is the relevant space of observables in the particle system. We call $\alpha_{\mathbb{S}}(N, n)$ the constant defined as in (6.5.1), with the infimum restricted to positive functions $f \in \mathbb{S}$.

Our main results for the GRT process are the following estimates independent of N.

Theorem 6.5.2. *Fix $n \in \mathbb{N}$ and assume that ν is strictly separating. Then there exists $\alpha(\nu) > 0$ such that for any $N \in \mathbb{N}$, $N \geq 2$,*

$$\alpha(N, n) \geq \alpha(\nu). \tag{6.5.2}$$

Moreover, if ν is the one-point crossover, then

$$\alpha(N, n) \geq \frac{1}{4(n+1)}\,, \tag{6.5.3}$$

and if ν is the uniform crossover, then

$$\alpha(N, n) \geq \frac{1}{4n}\,, \qquad \alpha_{\mathbb{S}}(N, n) \geq \frac{1}{2(n+2)}\,. \tag{6.5.4}$$

Remark 6.5.3. Consider the entropy production rate $\alpha(\Omega_{\varrho_N})$ for the process on Ω_{ϱ_N} associated to any admissible density sequence ϱ_N. This quantity can be defined as in (6.5.1) by restricting to invariant classes of functions with suitable symmetries. The estimates above then immediately provide the lower bound

$$\alpha(\Omega_{\varrho_N}) \geq \alpha(\nu)\,.$$

In particular, in the case of uniform crossover one finds $\alpha(\Omega_{\varrho_N}) \geq 1/4n$, for any admissible ϱ_N.

Remark 6.5.4. If the recombination measure is only assumed to be separating, then the GRT process may fail to be irreducible. In particular, some assumption such

as the strict separation defined above is necessary for the statement in Theorem 6.5.2. For an example of non irreducible process with separating ν consider $N = 2, n = 4$ and suppose $\nu(A) = \frac{1}{6}$ for all $A \subset [4]$ with $|A| = 2$. Clearly, ν is separating, but if we consider the initial configuration η with $\eta(1) = 0000, \eta(2) = 1111$, then the number of 1's in each particle remains even at all times.

For the proof of Theorem 6.5.2 we establish some new functional inequalities for permutations which imply a modified logarithmic Sobolev inequality for the GRT, see Section 7.3. Concerning upper bounds on the constant $\alpha(N, n)$ we establish an estimate valid for arbitrary ν, which essentially shows that $\alpha(N, n)$ cannot be larger than $4/n$ for n large, provided N is taken large enough, possibly depending on n.

Proposition 6.5.5. *For any $n \in \mathbb{N}$, any distribution ν on $[n]$,*

$$\limsup_{N\to\infty} \alpha(N, n) \leq \frac{4}{n} + O\left(\frac{1}{n^2}\right).$$

In this sense, the bounds in (6.5.3) and (6.5.4) can be considered to be optimal up to constants.

6.6 Kac's program completed

One of the main motivations behind Kac's program is the derivation of quantitative bounds on the speed of convergence to equilibrium for the nonlinear equation. In our setting, as a corollary of our analysis we obtain the following relative entropy estimates. We refer to [130, 31] for a discussion of related entropy decay estimates in the context of kinetic models. In particular, in our setup, one can say that Cercignani's conjecture holds true. See also [59] for related results in a discrete setting under positive curvature assumptions.

Theorem 6.6.1. *Assume that ν is strictly separating. For any $p \in \mathcal{P}(\Omega)$, let p_t denote the solution of* (6.1.1) *with $p_0 = p$ and let $\pi = \otimes_{i=1}^n p_i$ denote the associated equilibrium. Then for all $t \geq 0$,*

$$H(p_t \,|\, \pi) \leq e^{-\alpha(\nu)\, t} H(p \,|\, \pi), \tag{6.6.1}$$

where $\alpha(\nu) > 0$ is the constant in Theorem 6.5.2. In particular, $\alpha(\nu) \geq 1/4(n+1)$ for one-point crossover, and $\alpha(\nu) \geq 1/2(n+2)$ for uniform crossover.

It is interesting to note that the constant $\alpha(\nu)$ does not depend on the initial datum p in any way. We point out that, in the case of the one-point crossover and

uniform crossover, the above estimates were already obtained in [29] by direct analysis of the entropy production functional of the nonlinear equation, with a slightly better constant actually: $\alpha(\nu) \geq 1/(n+1)$ in both cases. Moreover, [29] also shows that the $1/n$ decay of the constant $\alpha(\nu)$ in these cases is optimal up to a constant independent of n. Besides extending the bounds of [29] to all strictly separating distribution ν, an interesting feature of Theorem 6.6.1 is that its proof takes a completely different route. Namely, it is based on the implementation in our setting of Kac's original idea. More precisely, (6.6.1) is derived from the uniform control on entropy production provided by Theorem 6.5.2, see also Remark 6.5.3, together with the approximation, as $N \to \infty$, of both $H(p \,|\, \pi)$ and $H(p_t \,|\, \pi)$ in terms of the corresponding entropies for the N-particle system.

We also obtain the following general bounds on the convergence to equilibrium for the nonlinear chain. Recall the definition (6.1.2) of the constant $r(\nu) \in (0,1)$.

Theorem 6.6.2. *Assume that ν is separating. For any $p \in \mathcal{P}(\Omega)$, let p_t denote the solution of* (6.1.1) *with $p_0 = p$ and let $\pi = \otimes_{i=1}^n p_i$ denote the associated equilibrium. Then for all $t \geq 0$,*

$$H(p_t \,|\, \pi) \leq \tfrac{1}{2}\, n(n-1)\, H(p \,|\, \pi)\, e^{-D(\nu)\, t}, \tag{6.6.2}$$

where $D(\nu) := 1 - r(\nu)$. Moreover, for the total variation distance we have

$$\|p_t - \pi\|_{\mathrm{TV}} \leq \tfrac{1}{4}\, n^2(n-1)\, \|p - \pi\|_{\mathrm{TV}}\, e^{-D(\nu)\, t}. \tag{6.6.3}$$

We note that an estimate similar to (6.6.3) was obtained in [115] for a discrete time version of the nonlinear process. To prove Theorem 6.6.2 we use a coupling argument similar to that of [115], together with an explicit construction of the continuous time solution p_t in terms of all possible collision histories, which goes back to the pioneering works of Wild [133] and McKean [81, 80], see also [30]. It is interesting to note that in the case of uniform crossover one has $r(\nu) = 1/2$ and thus (6.6.2) provides an exponential decay which is much faster, as n becomes large, than the one provided by (6.6.1). Moreover, as mentioned, the $1/n$ rate is known to be optimal up to a constant independent of n for the estimate (6.6.1). This mismatch can be explained by observing that, because of the possibly large prefactor, (6.6.3) only provides information about the large time behavior while (6.6.1) expresses a contraction property of the relative entropy at all times, and that some particular initial distributions p may have a slow start in the relative entropy decay; see Lemma 7.3.4 for a concrete example.

Finally, we remark that the rate of exponential decay $D(\nu) = 1 - r(\nu)$ in The-

orem 6.6.2 is optimal, in the sense that $t^{-1} \log \|p_t - \pi\|_{\mathrm{TV}}$, as $t \to \infty$, cannot be smaller than $-D(\nu)$, see Remark 7.2.4. We refer to [50] for a related result on the optimal rate of decay in the context of Kac model.

Chapter 7

Propagation of chaos and entropy production

In this chapter we prove the results stated in chapter 6. The organization goes as follows. In Section 7.1 we present the main preliminary facts concerning the local central limit theorem and its applications to the proof of chaos results. In Section 7.2 we prove the uniform in time propagation of chaos stated in Theorem 6.4.1. This section also contains the proof of Theorem 6.6.2. Section 7.3 is devoted to the proofs of Theorem 6.5.2, Proposition 6.5.5, and Theorem 6.6.1. In the appendix we give the detailed proof of the local central limit theorem statement used in the main text.

Acknowledgements

We would like to thank Arnaud Guillin, Cyril Labbé, Hubert Lacoin and Alistair Sinclair, for insightful conversations related to the results in this chapter.

7.1 Local Central Limit Theorem and Chaos

A probability measure $p \in \mathcal{P}(\Omega)$ induces a probability μ on $X := \{0,1\}^K$ where $K = \sum_{i=1}^n q_i$, via the map

$$\sigma \in \Omega \mapsto \xi_{i,x} = \mathbf{1}(\sigma_i = x)\,, \qquad i = 1, \dots, n\,;\ x \in \{1, \dots, q_i\}. \tag{7.1.1}$$

That is, μ is the push forward of p by the above map. Note that we did not include the indicator variable $\mathbf{1}(\sigma_i = 0)$ since this is uniquely determined as the indicator of the event $\xi_{i,x} = 0$ for all $x \in \{1, \dots, q_i\}$. When $q_i = 1$ for all i, then $\Omega = \{0,1\}^n$ can be identified with X, σ with ξ, and μ with p.

7.1.1 Central limit theorem

The next results are concerned with the behavior of the sum of independent copies $\xi(1), \ldots, \xi(N)$ of a random variable ξ with values in X and distribution $\mu \in \mathcal{P}(X)$:

$$S_N = \sum_{j=1}^{N} \xi(j).$$

Thus, S_N is a random vector in $\{0, \ldots, N\}^K$. We use the notation

$$\langle t, s \rangle = \sum_{i,x} t_{i,x} s_{i,x}$$

if t and s are indexed by $i = 1, \ldots, n$ and $x = 1, \ldots, q_i$. We call V_1 the covariance matrix of μ,

$$V_1(i, x; i', x') = \mu(\xi_{i,x}\xi_{i',x'}) - \mu(\xi_{i,x})\mu(\xi_{i',x'}), \qquad i = 1, \ldots, n\,,\ x = 1, \ldots, q_i$$

Thus V_1 is a symmetric nonnegative definite $K \times K$ matrix. If $\det V_1 \neq 0$ we say that μ is *nondegenerate*. The central limit theorem asserts that if μ is nondegenerate, then as $N \to \infty$ one has the weak convergence

$$\frac{1}{\sqrt{N}}\, V_1^{-1/2} \left(S_N - \mu^{\otimes N}(S_N)\right) \;\longrightarrow\; N(0, \mathbf{1}_K)\,, \tag{7.1.2}$$

where $\mu^{\otimes N}(S_N) \in [0, N]^K$ is the mean of the vector S_N under the product measure $\mu^{\otimes N}$, and $\mathbf{1}_K$ denotes the $K \times K$ identity matrix, so that $N(0, \mathbf{1}_K)$ is the standard normal in K dimensions. Note that when μ is induced by a measure $p \in \mathcal{P}(\Omega)$ as described in (7.1.1), then $\mu^{\otimes N}(S_N)_{i,x} = N\pi_{i,x}$ for all i, x, where $\pi_{i,x}$ are the marginals of p.

The statement (7.1.2) clearly requires that μ is nondegenerate. However, one can obtain similar statements in the case of degenerate measures, provided one reduces to the nondegenerate modes by eliminating the degenerate ones. More precisely, one can take the eigenvectors of V_1 with nonzero eigenvalues as the new variables. A simple example is obtained if e.g. $q_i \equiv 1$ and μ gives probability $1/2$ to all 1's and probability $1/2$ to all 0's. Here one simply removes all variables but one.

7.1.2 Local central limit theorem

We will need a local version of the central limit theorem. For this we assume the following stronger notion of nondegeneracy, which we refer to as irreducibility.

Definition 7.1.1. A measure $\mu \in \mathcal{P}(X)$ is called *irreducible* if for all $i = 1, \dots, n$, for all $x \in \{0, \dots, q_i\}$, there exists $\xi \in X$ such that $\mu(\xi)$ and $\mu(\xi^{(i,x)})$ are both positive, where $\xi^{(i,x)}$ denotes the vector ξ with the (i,x)-th coordinate flipped, that is $\xi^{(i,x)}_{j,y} = \xi_{j,y}$ for all $(j,y) \neq (i,x)$, and $\xi^{(i,x)}_{i,x} = 1 - \xi_{i,x}$.

It is immediate to check that if $p \in \mathcal{P}(\Omega)$ is irreducible in the sense of Definition 6.3.3 then the measure μ induced on X by p as in (7.1.1) is irreducible in the sense of Definition 7.1.1.

Proposition 7.1.2. *Suppose $\mu \in \mathcal{P}(X)$ is irreducible. Then there exists a finite constant $C = C(\mu)$ such that for all $N \in \mathbb{N}$,*

$$\max_{M_N} \left| \mu^{\otimes N}\left(S_N = M_N\right) \; - \; \frac{e^{-\frac{1}{2}\langle z_N, z_N\rangle}}{(2\pi N)^{K/2}\sqrt{\det V_1}} \right| \leq \frac{C}{N^{(K+1)/2}},$$

where

$$z_N := \frac{1}{\sqrt{N}}\, V_1^{-1/2}\left(M_N - \mu^{\otimes N}(S_N)\right),$$

and the maximum is over all possible values $M_N \in \{0, \dots, N\}^K$.

Noting that $\Omega_{\varrho_N} = \{S_N = M_N\}$ with $M_N = N\varrho_N$, and that in this case

$$\langle z_N, z_N\rangle = N\langle \varrho_N - \pi, V_1^{-1}(\varrho_N - \pi)\rangle,$$

the following is an immediate corollary of Proposition 7.1.2.

Corollary 7.1.3. *Suppose $\mu \in \mathcal{P}(X)$ is irreducible, and let ϱ_N be an admissible sequence such that*

$$\langle \varrho_N - \pi, \varrho_N - \pi\rangle = O(1/N), \tag{7.1.3}$$

where $\pi = \mu(\xi)$ is the vector of the expected values of μ. Then there exists a constant $c = c(\mu) > 0$ such that for N sufficiently large

$$\mu^{\otimes N}\left(\Omega_{\varrho_N}\right) \geq \frac{c}{N^{K/2}}.$$

In particular,

$$\lim_{N\to\infty} \frac{1}{N} \log \mu^{\otimes N}\left(\Omega_{\varrho_N}\right) = 0.$$

We note that the condition (7.1.3) corresponds to “normal" fluctuations

$$\langle z_N, z_N\rangle = O(1),$$

and that the corollary applies, in particular, to the canonical sequence $\varrho_N = \varrho^\pi$ defined in (6.2.4), since $\langle \varrho^\pi - \pi, \varrho^\pi - \pi \rangle = O(1/N^2)$ in that case.

The proof of Proposition 7.1.2 will be given in the appendix. Here we pause for some remarks on the assumptions we made, and then discuss the main applications to chaos.

Lemma 7.1.4. *If μ is irreducible then it is nondegenerate. The converse does not hold.*

Proof. If μ is degenerate, then for any fixed (i, x), the variable $\xi_{i,x}$ can be written μ-a.s. as a nontrivial linear combination of the other variables $\xi_{j,y}$, $(j, y) \neq (i, x)$. In particular, the value of $\xi_{i,x}$ is μ-a.s. determined by the other variables. But this is not possible if μ is irreducible since by assumption there is always at least one value of all the other variables for which both values $\xi_{i,x} = 0, 1$ happen with positive μ probability. This proves the first assertion. To violate the converse, consider the following example: $n = 3$, $q_i \equiv 1$, so that $X = \{0, 1\}^3$ and suppose that μ gives probability $1/4$ to the following four configurations $101, 110, 011, 000$, and probability 0 to the four remaining configurations. Then one checks that $V_1 = \frac{1}{4}\mathbf{1}_3$. In particular, μ is nondegenerate. However, μ is not irreducible since the condition in Definition 7.1.1 is violated at $i = 1$. □

Let us remark that some irreducibility assumption is necessary for the local CLT statement in Proposition 7.1.2. Consider the same counterexample from the proof of Lemma 7.1.4. In this case one checks easily that if the first component of S_N is even, then the sum of the remaining two components must be even as well. This shows that the event $S_N = M_N$ has probability zero for many admissible sequences such that Corollary 7.1.3 would predict $\mu^{\otimes N}(S_N = M_N) > 0$. Thus, Proposition 7.1.2 does not hold for all nondegenerate μ. The next lemma elucidates the role of the irreducibility assumption.

Lemma 7.1.5. *Suppose μ is irreducible. Then there exists a constant $c = c(\mu) > 0$ such that the characteristic function $\psi(t) = \mu(e^{i\langle t,\xi\rangle})$, $t \in \mathbb{R}^K$, satisfies*

$$|\psi(t)| \leq e^{-c\langle t,t\rangle}, \qquad \textit{for all } \ t \in [-\pi, \pi]^K.$$

Proof. We write

$$|\psi(t)|^2 = \left|\mu\left[e^{i\langle t,\xi\rangle}\right]\right|^2 = \mu\left[\cos\langle t, \xi\rangle\right]^2 + \mu\left[\sin\langle t, \xi\rangle\right]^2 = \sum_{\xi,\xi' \in X} \mu(\xi)\mu(\xi') \cos\langle t, \xi - \xi'\rangle,$$

where the last equation uses the identity $\cos(\alpha-\beta) = \sin(\alpha)\sin(\beta)+\cos(\alpha)\cos(\beta)$.

If $|\theta| \le \pi$, then $\cos(\theta) \le 1 - 2\theta^2/\pi^2$, and therefore,

$$\cos\langle t, \xi - \xi' \rangle \le \begin{cases} 1 - \frac{2\langle t, \xi - \xi' \rangle^2}{\pi^2} & \text{if } |\xi - \xi'|_1 = 1 \\ 1 & \text{if } |\xi - \xi'|_1 \ne 1 \end{cases}$$

where $|\xi - \xi'|_1 = \sum_{i,x} |\xi_{i,x} - \xi'_{i,x}|$. Since $|\xi - \xi'|_1 = 1$ iff $\xi' = \xi^{(i,x)}$ for some i, x,

$$|\psi(t)|^2 \le 1 - \frac{2}{\pi^2} \sum_{i,x} \sum_{\xi \in X} \mu(\xi)\mu(\xi^{(i,x)})\, t_{i,x}^2 \le 1 - 2c\langle t, t \rangle,$$

where

$$c := \frac{1}{\pi^2} \inf_{i,x} \sum_{\xi \in X} \mu(\xi)\mu(\xi^{(i,x)}).$$

The irreducibility of μ is equivalent to $c > 0$. Using $x \le e^{\frac{1}{2}(x^2-1)}$, $x \in [0,1]$, with $x = |\psi(t)|$, we conclude

$$|\psi(t)| \le e^{-c\,\langle t,t \rangle}.$$

□

We turn to the applications to Kac chaos and entropic chaos.

7.1.3 Kac chaos

Recall the definition of $\gamma(p, \varrho_N)$ and of the canonical tensor product $\gamma_N(p)$ in Definition 6.3.2.

Theorem 7.1.6. *Suppose $p \in \mathcal{P}(\Omega)$ is irreducible and let ϱ_N be an admissible sequence such that*

$$\langle \varrho_N - \pi, \varrho_N - \pi \rangle = O(1/N).$$

Then for all $k = 1, \dots, N$,

$$\|P_k \gamma(p, \varrho_N) - p^{\otimes k}\|_{\text{TV}} \le \frac{C\,k}{\sqrt{N}}, \tag{7.1.4}$$

for some constant $C = C(p)$. Moreover, when $\varrho_N = \varrho^\pi$, the canonical tensor product $\gamma_N(p)$ satisfies the stronger estimate

$$\|P_k \gamma_N(p) - p^{\otimes k}\|_{\text{TV}} \le \frac{C\,k}{N}. \tag{7.1.5}$$

Proof. We prove (7.1.4) first, and then show how to obtain (7.1.5). Let $\hat{\xi}_{i,x}(j) =$

$\xi_{i,x}(j) - (\varrho_N)_{i,x}$. We use the shorthand notation $\gamma_N = \gamma(p, \varrho_N)$ and $\mu_N = p^{\otimes N}$. For any $f : \Omega^N \mapsto \mathbb{R}$ we have

$$\gamma_N(f) - \mu_N(f) = \frac{\mu_N(f(\mathbf{1}_{\Omega_{\varrho_N}} - \mu_N(\Omega_{\varrho_N}))}{\mu_N(\Omega_{\varrho_N})},$$

Since $\Omega_{\varrho_N} = \{S_N = N\varrho_N\}$, using the Fourier transform we write

$$\mu_N(\Omega_{\varrho_N}) = \frac{1}{(2\pi)^K} \int_{[-\pi,\pi]^K} dt\, \mu_N\left(e^{i\langle t, \hat{S}_N\rangle}\right),$$

where $\hat{S}_N = \sum_{j=1}^N \hat{\xi}(j)$. Set $V_N := NV_1$. The product structure of μ_N and the change of variables $s = V_N^{1/2} t = \sqrt{N} V_1^{1/2} t$ imply

$$\mu_N(\Omega_{\varrho_N}) = \frac{1}{B_N(2\pi)^K} \int_{Q_{N,K}} ds\, \mu_N\left(e^{i\langle V_N^{-1/2} s, \hat{\xi}(1)\rangle}\right)^N, \tag{7.1.6}$$

where $Q_{N,K} = V_N^{1/2}[-\pi, \pi]^K$ and $B_N = \sqrt{\det V_N} = N^{K/2}\sqrt{\det V_1}$.

In the same way, for any $f = f(\xi(1), \ldots, \xi(k))$, we have

$$\mu_N(f\, \mathbf{1}_{\Omega_{\varrho_N}}) = \frac{1}{B_N(2\pi)^K} \int_{Q_{N,K}} ds\, \mu_N\left(e^{i\langle V_N^{-1/2} s, \hat{\xi}(1)\rangle}\right)^{N-k} \mu_N\left(f\, e^{i\langle V_N^{-1/2} s, \hat{S}_k\rangle}\right).$$

In conclusion, we have

$$\gamma_N(f) - \mu_N(f) = \frac{\int_{Q_{N,K}} ds\, \psi_N(s)^{N-k} \mu_N\left(f;\, e^{i\langle V_N^{-1/2} s, \hat{S}_k\rangle}\right)}{\int_{Q_{N,K}} ds\, \psi_N(s)^N}, \tag{7.1.7}$$

where

$$\psi_N(s) = \mu_N\left(e^{i\langle V_N^{-1/2} s, \hat{\xi}(1)\rangle}\right),$$

and we use the notation $\mu_N(f; g) = \mu_N(fg) - \mu_N(f)\mu_N(g)$ for the covariance of f, g. From Corollary 7.1.3 we known that (7.1.6) is at least $cN^{-K/2}$, and thus the denominator in (7.1.7) is at least some constant $c' > 0$. Therefore, it suffices to show that the numerator is bounded by

$$\int_{Q_{N,K}} ds\, |\psi_N(s)|^{N-k} \left|\mu_N\left(f;\, e^{i\langle V_N^{-1/2} s, \hat{S}_k\rangle}\right)\right| \le C\, |f|_\infty \frac{k}{\sqrt{N}}.$$

From Lemma 7.1.5 we know that $|\psi_N(s)| \le e^{-a\langle s,s\rangle/N}$ for some constant $a = a(p) > 0$. Notice that we can assume without loss of generality that $k \le N/2$, since otherwise the result (7.1.4) is trivial. Thus $|\psi_N(s)|^{N-k} \le e^{-a\langle s,s\rangle/2}$ and it is sufficient

to show that

$$\left|\mu_N\left(f;\, e^{i\langle V_N^{-1/2}s,\hat{S}_k\rangle}\right)\right| \le C\,|f|_\infty\,\frac{k}{\sqrt{N}}\,\langle s,s\rangle, \tag{7.1.8}$$

for all $s \in Q_{N,K}$. Recalling that $|e^{i\theta}-1| \le |\theta|$, $\theta \in \mathbb{R}$, and using Schwarz' inequality,

$$|\mu_N\left(f;\, e^{i\langle V_N^{-1/2}s,\hat{S}_k\rangle}\right)| \le \frac{1}{\sqrt{N}}\,|f|_\infty\,\mu_N\left(\langle V_1^{-1/2}s,\hat{S}_k\rangle^2\right).$$

Now we observe that

$$\begin{aligned}\mu_N\left(\langle V_1^{-1/2}s,\hat{S}_k\rangle^2\right) &= k^2\langle V_1^{-1/2}s,\pi-\varrho_N\rangle^2 + k\langle s,s\rangle\\ &\le k^2\langle s,V_1^{-1}s\rangle\langle\pi-\varrho_N,\pi-\varrho_N\rangle + k\langle s,s\rangle \le Ck\langle s,s\rangle,\end{aligned}$$

where we use $\mu_N(\hat{S}_k) = k\,(\pi-\varrho_N)$, the independence of the $\hat{\xi}(j)$, and $\langle\pi-\varrho_N,\pi-\varrho_N\rangle \le C/k$ which follows from the assumption $\langle\pi-\varrho_N,\pi-\varrho_N\rangle = O(1/N)$. This proves (7.1.4).

To prove (7.1.5), note that it is sufficient to prove (7.1.8) with $\sqrt{N}$ replaced by N in the right hand side. For this, we are going to use the fact that $\langle\pi-\varrho_N,\pi-\varrho_N\rangle = O(1/N^2)$ when $\varrho_N = \varrho^\pi$, see (6.2.4). Let us first consider the function $\tilde{f} = f - g$, where

$$g(\xi(1),\dots,\xi(k)) = \frac{1}{k}\langle V_1^{-1}v,\hat{S}_k\rangle, \qquad v = \mu_N(f;\hat{S}_k).$$

The function g can be seen as a linear approximation of f. Notice that

$$\mu_N\left(f-g\;;\;\langle V_N^{-1/2}s,\hat{S}_k\rangle\right) = 0. \tag{7.1.9}$$

Indeed, by independence $\mu_N((\hat{S}_k)_{i,x};(\hat{S}_k)_{j,y}) = kV_1(i,x;j,y)$, and therefore for all s,

$$\mu_N\left(g\,;\,\langle V_N^{-1/2}s,\hat{S}_k\rangle\right) = \langle v,V_N^{-1/2}s\rangle = \mu_N\left(f\,;\,\langle V_N^{-1/2}s,\hat{S}_k\rangle\right).$$

Recalling that $|e^{i\theta}-1-i\theta| = |R(\theta)| \le \frac{1}{2}\theta^2$, $\theta\in\mathbb{R}$, from (7.1.9) we have

$$\left|\mu_N\left(f-g\,;\,e^{i\langle V_N^{-1/2}s,\hat{S}_k\rangle}\right)\right|^2 \le \frac{1}{N^2}\,(\mathrm{Var}(f)+\mathrm{Var}(g))\,\mu_N\left(\langle V_1^{-1/2}s,\hat{S}_k\rangle^4\right),$$

where we use the inequality $|\mu_N(f-g;R)|^2 \le 2\,(\mathrm{Var}(f)+\mathrm{Var}(g))\,\mu_N(R^2)$, and we use Var for the variance w.r.t. μ_N. Next, observe that $\mathrm{Var}(f) \le |f|_\infty^2$ and

$$\mathrm{Var}(g) = \frac{1}{k}\langle V^{-1}v,v\rangle \le C\,\mathrm{Var}(f) \le C|f|_\infty^2.$$

We are going to show that

$$\mu_N\left(\langle V_1^{-1/2}s, \hat{S}_k\rangle^4\right) \leq Ck^2\langle s, s\rangle^2\,. \tag{7.1.10}$$

Suppose for a moment that (7.1.10) holds. Then we conclude that

$$\left|\mu_N\left(f-g\,;\, e^{i\langle V_N^{-1/2}s, \hat{S}_k\rangle}\right)\right| \leq C\,|f|_\infty\,\frac{k}{N}\,\langle s, s\rangle.$$

This proves $|\gamma_N(\tilde{f}) - \mu_N(\tilde{f})| \leq C\,|f|_\infty\, k/N$. However, noting that $\gamma_N(g) = 0$, we have

$$|\gamma_N(f) - \mu_N(f)| \leq |\gamma_N(\tilde{f}) - \mu_N(\tilde{f})| + |\mu_N(g)|.$$

The desired conclusion $|\gamma_N(f) - \mu_N(f)| \leq C\,|f|_\infty\, k/N$ then follows from

$$|\mu_N(g)| \leq |\langle V_1^{-1}v, \pi - \varrho^\pi\rangle| \leq \langle V_1^{-1}v, V_1^{-1}v\rangle^{1/2}\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle^{1/2} \leq \frac{C|f|_\infty k}{N},$$

where we use $\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle = O(1/N^2)$, and $\langle V_1^{-1}v, V_1^{-1}v\rangle \leq Ck|f|_\infty$.

Thus, it remains to prove (7.1.10). Notice that it is sufficient to prove

$$\mu_N\left(\langle \hat{S}_k, \hat{S}_k\rangle^2\right) \leq Ck^2\,. \tag{7.1.11}$$

We write $\tilde{\xi} = \hat{\xi} - \mu(\hat{\xi})$ and $\tilde{S}_k = \hat{S}_k - \mu_N(\hat{S}_k)$ for the corresponding sums. Then one checks that

$$\begin{aligned}\langle \hat{S}_k, \hat{S}_k\rangle^2 \leq\; & \langle \tilde{S}_k, \tilde{S}_k\rangle^2 + Ck\langle \tilde{S}_k, \tilde{S}_k\rangle^{3/2}\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle^{1/2} \\ & + Ck^2\langle \tilde{S}_k, \tilde{S}_k\rangle\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle + Ck^3\langle \tilde{S}_k, \tilde{S}_k\rangle^{1/2}\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle^{3/2} \\ & + Ck^4\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle^2\,.\end{aligned}$$

Since $\langle \pi - \varrho^\pi, \pi - \varrho^\pi\rangle = O(1/N^2)$ we can restrict to prove (7.1.11) for $\tilde{S}_k$ instead of $\hat{S}_k$. Since $\tilde{\xi}$ are centered the estimate follows easily by expanding $\langle \tilde{S}_k, \tilde{S}_k\rangle^2$ and observing that the dominant terms are of the form $\langle \tilde{\xi}(i), \tilde{\xi}(j)\rangle^2$, and their contribution is of order k^2. □

We notice that the trick of replacing f by $f - g$ in the proof of Theorem 7.1.6 allowed us to obtain the decay rate $O(1/N)$ instead of $O(1/\sqrt{N})$. This idea was used in [24] for the proof of a related "equivalence of ensembles" result.

Corollary 7.1.7. *For any irreducible $p \in \mathcal{P}(\Omega)$, if $\eta \in \Omega^N$ is distributed according to the canonical tensor product $\gamma_N(p)$, letting $\lambda_\eta = \frac{1}{N}\sum_{j=1}^N \delta_{\eta(j)}$ denote the corresponding*

empirical measure,

$$\mathbb{E}\Big[\|p-\lambda_\eta\|_{\mathrm{TV}}\Big] \leq \frac{C}{\sqrt{N}}, \tag{7.1.12}$$

for some constant $C = C(p)$. Moreover, the same estimate holds with $\sqrt{N}$ replaced by $N^{1/4}$ if η is distributed according to $\gamma(p, \varrho_N)$, for any admissible sequence ϱ_N satisfying (7.1.3).

Proof. We write

$$\begin{aligned}\mathbb{E}\left[\|p-\lambda_\eta\|_{\mathrm{TV}}\right] &= \frac{1}{2}\sum_{\sigma\in\Omega}\mathbb{E}\left[|\lambda_\eta(\sigma)-p(\sigma)|\right] \\ &\leq \frac{1}{2}\left(\sum_{\sigma\in\Omega} p(\sigma)\mathbb{E}\left[(h_\eta(\sigma)-1)^2\right]\right)^{1/2}, \end{aligned}\tag{7.1.13}$$

where $h_\eta(\sigma) = \frac{\lambda_\eta(\sigma)}{p(\sigma)}$ and we have used Schwarz' inequality for the product measure $p \times \mathbb{E}$. Now,

$$\begin{aligned}\sum_\sigma p(\sigma)\mathbb{E}\left[(h_\eta(\sigma)-1)^2\right] &= -1 + \frac{1}{N}\,\mathbb{E}\left[\frac{1}{p(\eta(1))}\right] \\ &\quad + \frac{N(N-1)}{N^2}\,\mathbb{E}\left[\frac{\mathbf{1}_{\eta(1)=\eta(2)}}{p(\eta(1))}\right],\end{aligned}\tag{7.1.14}$$

where we use

$$\mathbb{E}\left[\frac{1}{p(\eta(1))}\right] = \sum_\sigma \frac{\mathbb{P}(\eta(1)=\sigma)}{p(\sigma)}, \quad \mathbb{E}\left[\frac{\mathbf{1}_{\eta(1)=\eta(2)}}{p(\eta(1))}\right] = \sum_\sigma \frac{\mathbb{P}(\eta(1)=\eta(2)=\sigma)}{p(\sigma)}.$$

Since $\mathbb{E}\left[1/p(\eta(1))\right] \leq 1/p_*$, where $p_* = \min_{\sigma:\, p(\sigma)>0} p(\sigma)$ we see that the second term in (7.1.14) is bounded by $1/(p_* N)$. Next, consider the function $f(\eta(1), \eta(2)) = \mathbf{1}_{\eta(1)=\eta(2)}/p(\eta(1))$. Note that $p^{\otimes 2}(f) = 1$. Then, Theorem 7.1.6 implies

$$\left|-1 + \sum_\sigma \frac{\mathbb{P}(\eta(1)=\eta(2)=\sigma)}{p(\sigma)}\right| \leq \frac{C}{p_* N}.$$

Since $N(N-1)/N^2 = 1-1/N$, we have shown that (7.1.14) is bounded by C/N for some new constant C depending only on p. Together with (7.1.13) this concludes the proof of (7.1.12). Finally, if instead η is distributed according to $\gamma(p, \varrho_N)$, for an arbitrary admissible sequence ϱ_N satisfying (7.1.3) we may repeat all the steps above and use the first part of Theorem 7.1.6 to conclude that (7.1.14) this time is bounded by $C/\sqrt{N}$, which implies the claimed bound with $N^{1/4}$ in place of $\sqrt{N}$. □

7.1.4 Entropic chaos and Fisher chaos

The next result shows how to use the local CLT to obtain convergence of the relative entropy of tensor products. Following [31] we refer to this as entropic chaos.

Proposition 7.1.8. *Suppose $p \in \mathcal{P}(\Omega)$ is irreducible and let ϱ_N be an admissible sequence such that*

$$\langle \varrho_N - \pi, \varrho_N - \pi \rangle = O(1/N).$$

Then

$$\lim_{N\to\infty} \frac{1}{N} H_N(\gamma(p, \varrho_N) \,|\, \gamma(\pi, \varrho_N)) = H(p \,|\, \pi).$$

Proof. We write

$$H_N(\gamma(p, \varrho_N) \,|\, \gamma(\pi, \varrho_N)) = \gamma(p, \varrho_N) \left[\log \left(\frac{p^{\otimes N}}{\pi^{\otimes N}} \right) \right] + \log \left(\frac{\pi^{\otimes N}(\Omega_{\varrho_N})}{p^{\otimes N}(\Omega_{\varrho_N})} \right). \tag{7.1.15}$$

From Corollary 7.1.3 we obtain

$$\lim_{N\to\infty} \frac{1}{N} \log \left(\frac{\pi^{\otimes N}(\Omega_{\varrho_N})}{p^{\otimes N}(\Omega_{\varrho_N})} \right) = 0.$$

By symmetry, the first term in (7.1.15) equals

$$N\, P_1\gamma(p, \varrho_N) \left[\log \left(\frac{p}{\pi} \right) \right].$$

Therefore the result follows from Theorem 7.1.6. □

Another consequence of the local CLT is the following upper semi-continuity property, see [31] for a similar statement in the kinetic setting.

Proposition 7.1.9. *For each N, let $\mu^{(N)}$ be a symmetric probability on Ω_{ϱ_N} and let μ_k be a probability Ω^k such that*

$$P_k\mu^{(N)} \longrightarrow \mu_k$$

weakly for some integer k. Then, for any admissible sequence satisfying (7.1.3)*,*

$$\frac{H(\mu_k \,|\, \pi^{\otimes k})}{k} \leq \liminf_{N\to\infty} \frac{H_N(\mu^{(N)} | \gamma(\pi, \varrho_N))}{N}.$$

Proof. An application of the Shearer inequality (2.4.5) with respect to the measure $\pi^{\otimes N}$ yields

$$\sum_{A\in\mathcal{A}} H\left(P_A\mu^{(N)}|P_A\pi^{\otimes N}\right) \leq n_+(\mathcal{A}) H\left(\mu^{(N)}\,|\,\pi^{\otimes N}\right),$$

where $\mathcal{A}$ is any family of sets covering $[N] = \{1,\dots,N\}$, $n_+(\mathcal{A}) = \max_j \#\{A \in \mathcal{A} : A \ni j\}$, and $P_A\nu$ denotes the marginal on the variables $\{\eta(j),\, j \in A\}$ of a probability $\nu \in \mathcal{P}(\Omega^N)$. If we take $\mathcal{A} = \{A \subset [N] : |A| = k\}$, then $n_+(\mathcal{A}) = \binom{N-1}{k-1}$. Moreover, by symmetry $P_A\mu^{(N)} = P_k\mu^{(N)}$ for all $A \in \mathcal{A}$. Since $\binom{N}{k}/\binom{N-1}{k-1} = N/k$, this proves

$$H\left(P_k\mu^{(N)}\,|\,\pi^{\otimes k}\right) \leq \frac{k}{N} H\left(\mu^{(N)}\,|\,\pi^{\otimes N}\right). \tag{7.1.16}$$

See also Lemma 3.9 in [105], where (7.1.16) was derived with $\frac{k}{N}$ replaced by $\frac{k}{N}(1+O(\frac{k}{N}))$ in the right hand side. On the other hand,

$$H_N\left(\mu^{(N)}\,|\,\pi^{\otimes N}\right) = H_N(\mu^{(N)}\,|\,\gamma(\pi,\varrho_N)) - \log\left(\pi^{\otimes N}(\Omega_{\varrho_N})\right).$$

The left hand side of (7.1.16) converges by assumption to $H\left(\mu_k\,|\,\pi^{\otimes k}\right)$. The desired conclusion then follows from Corollary 7.1.3. □

Remark 7.1.10. Both Proposition 7.1.8 and Proposition 7.1.9 hold with π replaced by any irreducible $p' \in \mathcal{P}(\Omega)$ with the same marginals as π.

One can also establish the following analogue of the Fisher chaos property discussed in [75]. Observe that if $\mu_{N,t} = \mu_N e^{t\mathcal{L}_N}$, for some $\mu_N \in \mathcal{P}(\Omega_{\varrho_N})$, then

$$\frac{d}{dt} H\left(\mu_{N,t}|\gamma(\pi,\varrho_N)\right)_{|_{t=0^+}} = D_N(f_N),$$

where $f_N := \mu_N/\gamma(\pi,\varrho_N)$, and

$$D_N(f_N) = \frac{1}{2N}\sum_{j<l}\sum_A \nu(A)\,\gamma(\pi,\varrho_N)\left[(f_N^{j,l,A} - f_N)\log\frac{f_N^{j,l,A}}{f_N}\right].$$

Here we use the notation $f_N^{j,l,A}(\eta) = f_N(\eta^{j,l,A})$, where $\eta^{j,l,A}$ is defined in (6.2.2). On the other hand, for the nonlinear equation (6.1.1) one has

$$\frac{d}{dt} H\left(p_t\,|\,\pi\right)_{|_{t=0^+}} = D_\pi(f),$$

where $f := p/\pi$, and

$$D_\pi(f) = \sum_A \nu(A)\, \pi \left[(f^A f^{A^c} - f) \log \frac{f^A f^{A^c}}{f} \right],$$

where $f^A(\sigma) = p_A(\sigma)/\pi_A(\sigma)$ is the density of the marginal on A of p with respect to π.

Proposition 7.1.11. *Under the same assumptions of Proposition 7.1.8,*

$$\lim_{N\to\infty} \frac{D_N(f_N)}{N} = D_\pi(f),$$

where $f_N := \gamma(p, \varrho_N)/\gamma(\pi, \varrho_N)$, *and* $f := p/\pi$.

Proof. Since

$$D_N(f_N) = -\frac{1}{N} \sum_{j<l} \sum_A \nu(A)\, \gamma(\pi, \varrho_N) \left[(f_N^{j,l,A} - f_N) \log f_N \right].$$

and

$$D_\pi(f) = -2 \sum_A \nu(A)\, \pi \left[(f^A f^{A^c} - f) \log f \right],$$

it suffices to show that for every $j, l \in [N]$ and $A \subset [n]$ one has

$$\begin{aligned} \lim_{N\to+\infty} \sum_{\eta\in\Omega_{\varrho_N}} & \gamma(\pi, \varrho_N)(\eta) \left(f_N(\eta^{j,l,A}) - f_N(\eta) \right) \log \left(f_N(\eta) \right) \\ & = 2 \sum_{\sigma\in\Omega} (p_A(\sigma_A) p_{A^c}(\sigma_{A^c}) - p(\sigma)) \log \left(\frac{p(\sigma)}{\pi(\sigma)} \right). \end{aligned}$$

Note that

$$\begin{aligned} & \sum_{\eta\in\Omega_{\varrho_N}} \pi_N(\eta) \left(f_N(\eta_{j,l,A}) - f_N(\eta) \right) \log \left(f_N(\eta) \right) \\ & = \sum_{\eta\in\Omega_{\varrho_N}} \frac{\prod_{\ell\neq j,l} p(\eta(l))}{p^{\otimes N}\left(\Omega_{\varrho_N}\right)} \left(p(\eta^{j,l,A}(j)) p(\eta^{j,l,A}(l)) - p(\eta(l)) p(\eta(j)) \right) \times \\ & \qquad \times \log \prod_{r=1}^N \frac{p(\eta(r))}{\pi(\eta(r))} \\ & = \sum_{\eta(l),\eta(j)} \frac{p^{\otimes(N-2)}\left(\Omega_{\varrho_N}^{\eta(j),\eta(l)}\right)}{p^{\otimes N}\left(\Omega_{\varrho_N}\right)} \left(p(\eta^{j,l,A}(j)) p(\eta^{j,l,A}(l)) - p(\eta(l)) p(\eta(j)) \right) \times \\ & \qquad \times \log \frac{p(\eta(l)) p(\eta(j))}{\pi(\eta(l)) \pi(\eta(j))}, \end{aligned}$$

where $\Omega_{\varrho_N}^{\eta(j),\eta(l)}$ is the event

$$\sum_{\ell\neq j,l} \mathbf{1}(\eta_i(\ell)=x) = (N-2)(\widetilde{\varrho_N})_{i,x}\,, \quad \forall i,x$$

and, for any fixed $\eta(j),\eta(l)$, $\widetilde{\varrho_N}$ denotes the density

$$(\widetilde{\varrho_N})_{i,x} = \frac{N}{N-2}\,(\varrho_N)_{i,x} - \frac{1}{N(N-2)}\,(\mathbf{1}(\eta_i(j)=x)+\mathbf{1}(\eta_i(l)=x))\,.$$

Therefore, it is sufficient to show that

$$\frac{p^{\otimes(N-2)}\left(\Omega_{\varrho_N}^{\eta(j),\eta(l)}\right)}{p^{\otimes N}\left(\Omega_{\varrho_N}\right)} \;\to\; 1, \tag{7.1.17}$$

for all $i,j\in[N]$, for all fixed values of $\eta(j),\eta(l)\in\Omega$. We note that both $\varrho_N, \widetilde{\varrho_N}$ satisfy condition (7.1.3). Moreover, using $\widetilde{\varrho}_N = (1+O(1/N))\varrho_N + O(1/N^2)$ we see that

$$z_N = \sqrt{N}\;V_1^{-1/2}\,(\varrho_N-\pi)\,, \quad \widetilde{z}_N = \sqrt{N}\;V_1^{-1/2}\,(\widetilde{\varrho}_N-\pi)$$

satisfy

$$\langle z_N, z_N\rangle - \langle \widetilde{z}_N, \widetilde{z}_N\rangle \to 0\,, \quad N\to\infty,$$

for all fixed values of $\eta(j),\eta(l)$. The desired claim (7.1.17) then follows from Proposition 7.1.2. □

7.2 Uniform in time propagation of chaos

The main goal in this section is to prove Theorem 6.4.1. We first introduce some notation. Let $\lambda_\eta\in\mathcal{P}(\Omega)$ be the empirical measure $\lambda_\eta := \frac{1}{N}\sum_{i=1}^N \delta_{\eta(i)}$, where $\eta := (\eta(1),\dots,\eta(N))\in\Omega^N$. We consider the Wasserstein distance on $\mathcal{P}(\Omega)$ associated to the Hamming distance on Ω:

$$W(p,q) := \inf_{\Gamma\in\Pi(p,q)} \sum_{\sigma,\sigma'\in\Omega} \Gamma(\sigma,\sigma') \sum_{i=1}^n \mathbf{1}_{\sigma_i\neq\sigma'_i},$$

where $p,q\in\mathcal{P}(\Omega)$, and $\Pi(p,q)$ denotes the set of all couplings of p and q. We remark that the total variation distance $\|p-q\|_{\mathrm{TV}}$ is defined as above with $\sum_{i=1}^n \mathbf{1}_{\sigma_i\neq\sigma'_i}$ replaced by $\mathbf{1}_{\sigma\neq\sigma'}$ and thus one has

$$\|p-q\|_{\mathrm{TV}} \le W(p,q) \le n\,\|p-q\|_{\mathrm{TV}}. \tag{7.2.1}$$

In contrast with the total variation distance, the distance W has a convenient monotonicity along the evolution, see Lemma 7.2.8 and Remark 7.2.9 below. Recall the definition of the non-separation probability $r(\nu)$ from (6.1.2).

Theorem 7.2.1. *Assume that $\mu_N \in \mathcal{P}(\Omega^N)$ is $p-$chaotic and the measure ν is separating. Then for all $k \in \mathbb{N}$ and any function of the form $\varphi_k := \varphi^1 \otimes \cdots \otimes \varphi^k : \Omega^k \mapsto \mathbb{R}$, where $\varphi^i : \Omega \mapsto \mathbb{R}$, and such that $\|\varphi_k\|_\infty \leq 1$, the following inequality holds*

$$|P_k \mu_{N,t}(\varphi_k) - p_t^{\otimes k}(\varphi_k)| \leq \frac{2k(k-1)}{N} + \frac{k^2 n^5}{D(\nu)\, N}(1 - e^{-D(\nu)\, t}) + 2k\, \mu_N \left[W\left(p, \lambda_{\chi_N}\right) \right],$$

where $\chi_N \in \Omega^N$ has distribution μ_N, and $D(\nu) := 1 - r(\nu)$.

The proof of Theorem 7.2.1 follows the steps of a general approach, the so called "abstract theorem", see [69, 106, 107], see also [37] for a review. In our discrete setting many aspects of this approach take a simpler form, and no extra assumption on the initial distribution is needed besides the p-chaoticity for some $p \in \mathcal{P}(\Omega)$. This general approach requires however several model-specific inputs. Our main original contribution here consists in establishing the key estimates stated in Theorem 7.2.2 and Theorem 7.2.5 below, where we prove new contraction inequalities for both the non linear and the linearized evolutions associated to (6.1.1). Before proving Theorem 7.2.1, let us show that it implies Theorem 6.4.1.

7.2.1 Proof of Theorem 6.4.1

It is well known that

$$\mu_N \left[W\left(p, \lambda_{\chi_N}\right) \right] \to 0$$

if and only if μ_N is p-chaotic; see e.g. Lemma 3.34 in [37]. This, combined with the estimate from Theorem 7.2.1, proves the uniform in time propagation of chaos asserted in Theorem 6.4.1. To prove the quantitative statement (6.4.1), it suffices to show that when $\mu_N = \gamma_N(p, \varrho_N)$, then

$$\mu_N \left[W\left(p, \lambda_{\chi_N}\right) \right] \leq \frac{C_1}{\sqrt{N}},$$

where C_1 is a constant depending on p and n. This statement follows from Corollary 7.1.7 by noting that $W(p, q) \leq n\|p - q\|_{\mathrm{TV}}$, for any $p, q \in \mathcal{P}(\Omega)$, see (7.2.1). We remark that as in Corollary 7.1.7 one has the same estimate with $\sqrt{N}$ replaced by $N^{1/4}$ if instead of $\gamma_N(p)$ we take $\gamma(p, \varrho_N)$ with an arbitrary admissible ϱ_N satisfying (7.1.3).

We turn to the proof of Theorem 7.2.1. We start with some preliminary facts.

7.2.2 Wild sums and McKean trees

Given $f, g : \Omega \to \mathbb{R}$, we adopt the notation

$$(f \circ g)_A := \frac{1}{2}\left(f_A \otimes g_{A^c} + g_A \otimes f_{A^c}\right), \qquad (7.2.2)$$
$$f \circ g := \sum_{A \subset [n]} \nu(A)(f \circ g)_A,$$

where $f_A(\sigma_A) := \sum_{\sigma_{A^c}} f(\sigma_A \sigma_{A^c})$. Then the nonlinear equation (6.1.1) is given by

$$\frac{d}{dt} p_t = p_t \circ p_t - p_t. \qquad (7.2.3)$$

The convolution product defined by $f \circ g$ is commutative and distributive, but not associative. Following Wild's original construction [133] we write the solution of (7.2.3) with initial datum $p_0 = p$, as

$$p_t = e^{-t} \sum_{k=1}^{\infty} (1 - e^{-t})^{k-1} p^{(k)}, \qquad (7.2.4)$$

where the $\{p^{(k)}\}_{k \geq 1}$ are probability measures on Ω defined inductively by

$$p^{(1)} = p, \qquad p^{(k)} = \frac{1}{k-1} \sum_{j=1}^{k-1} p^{(j)} \circ p^{(k-j)}, \quad k \geq 2. \qquad (7.2.5)$$

The validity of (7.2.4) can be easily checked by direct inspection, see e.g. the argument after (7.2.32) below for a similar computation.

Moreover, following McKean [81], we may express $p^{(k)}$ as a weighted sum over rooted binary trees γ with k leaves. We now recall the details of this construction, and refer the reader to [30] for further background. Let $\Gamma(k)$ denote the set of rooted binary trees with k leaves and call $\alpha_k(\gamma)$, $\gamma \in \Gamma(k)$, the probability over $\Gamma(k)$ obtained by the following procedure. $\Gamma(1) = \{\gamma_1\}$ is just the empty tree with only the root with $\alpha_1(\gamma_1) = 1$, and $\Gamma(2) = \{\gamma_2\}$ where γ_2 is the unique tree obtained by adding two children to the root, with $\alpha_2(\gamma_2) = 1$. Then, recursively, for any $\gamma_{k-1} \in \Gamma(k-1)$, consider all possible trees $\gamma_k^i \in \Gamma(k)$, $i = 1, \ldots, k-1$, obtained by adding two children to the i-th leaf of γ_{k-1}, and set $\alpha_k(\gamma_k^i) = \frac{1}{k-1}\alpha_{k-1}(\gamma_{k-1})$ for each one of them. This defines the weights $\alpha_k(\gamma)$ for any $\gamma \in \Gamma(k)$ and one checks, recursively, that $\sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) = 1$ for all $k \geq 1$ and that for all $k \geq 2$,

$$\alpha_k(\gamma) = \frac{1}{k-1} \alpha_j(\gamma_l) \alpha_{k-j}(\gamma_r), \qquad (7.2.6)$$

where γ_l and γ_r denote respectively the subtree of γ rooted at the left child of the

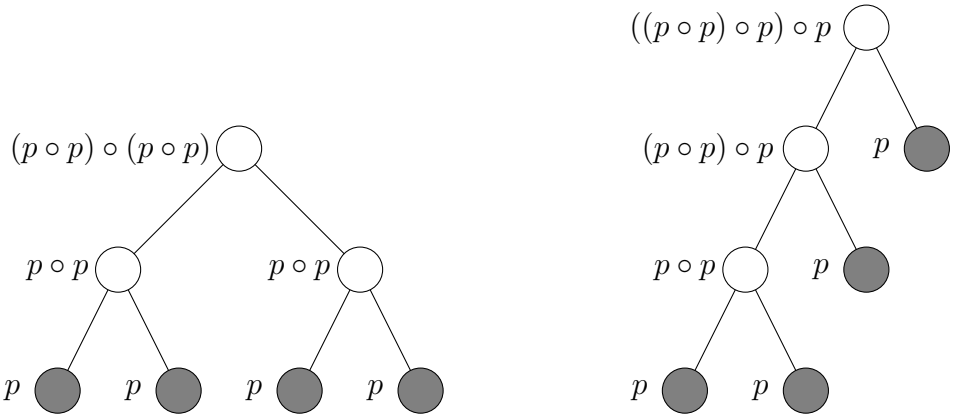

Figure 7.2.1: Two possible trees $\gamma, \gamma' \in \Gamma(4)$, and the corresponding distributions $C_\gamma(p) = (p \circ p) \circ (p \circ p)$ and $C_{\gamma'}(p) = ((p \circ p) \circ p) \circ p$.

root and the subtree of γ rooted at the right child of the root, while j denotes the number of leaves in γ_l. Then, by induction over k it follows that for all $k \in \mathbb{N}$,

$$p^{(k)} = \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) C_\gamma(p), \tag{7.2.7}$$

where $C_\gamma(p) \in \mathcal{P}(\Omega)$ is described as follows. Each internal node of γ represents a collision and the tree γ describes the collision history. Then $C_\gamma(p)$ represents the distribution obtained at the root after all collisions from γ have been performed, starting with the distribution p at each leaf of γ.

For example, if $k = 3$, there are only two trees $\gamma, \gamma' \in \Gamma(3)$, with $\alpha_3(\gamma) = \alpha_3(\gamma') = 1/2$ and

$$C_\gamma(p) = (p \circ p) \circ p = p \circ (p \circ p) = C_{\gamma'}(p). \tag{7.2.8}$$

We refer to Figure 7.2.1 for two examples of $C_\gamma(p)$ for $\gamma \in \Gamma(4)$.

Since each collision is of the form $q \circ q'$ for some $q, q' \in \mathcal{P}(\Omega)$ and since by (7.2.2) one has

$$p \circ q = \sum_{A \subset [n]} \bar{\nu}(A)\, p_A \otimes q_{A^c},$$

where $\bar{\nu}(A) = \frac{1}{2}(\nu(A) + \nu(A^c))$, one can write the following expansion for the resulting measure at the root:

$$C_\gamma(p) = \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A}) C_\gamma^{\vec{A}}(p), \tag{7.2.9}$$

where $\mathcal{V}_n$ is the set of subsets of $[n]$, so that $\vec{A} \in \mathcal{V}_n^{k-1}$ represents a pattern $\vec{A} =$

$(A_1, \dots, A_{k-1})$,

$$\nu(\vec{A}) = \prod_{i=1}^{k-1} \bar{\nu}(A_i) \tag{7.2.10}$$

and $C_\gamma^{\vec{A}}(p) \in \mathcal{P}(\Omega)$ is the distribution computed as follows. Each internal node v_i, $i = 1, \dots, k-1$, in γ (in some fixed order) is associated with the set A_i, and we attach the mark A_i to the edge connecting v_i to its left child and the mark A_i^c to the edge connecting v_i to its right child in γ. In this way we obtain a tree γ with marks on all its edges. Let d_i denote the depth of the i-th leaf of γ (e.g. counting from the leftmost leaf), and consider $p_{V_i(\vec{A})}$, the marginal of the measure p on the subset

$$V_i(\vec{A}) := \cap_{j=1}^{d_i} A_j^i, \tag{7.2.11}$$

where $A_1^i, \dots, A_{d_i}^i$ are the marks encountered along the edges of the unique path from the root to the i-th leaf. Then $C_\gamma^{\vec{A}}(p)$ is given by

$$C_\gamma^{\vec{A}}(p) = p_{V_1(\vec{A})} \otimes \cdots \otimes p_{V_k(\vec{A})}. \tag{7.2.12}$$

Note that some of the $V_i(\vec{A})$ may be empty. However, they form a partition of $[n]$, namely $V_i(\vec{A}) \cap V_j(\vec{A}) = \emptyset$ for $i \neq j$ and $\cup_{i=1}^k V_i(\vec{A}) = [n]$, which can be seen as the result of a fragmentation process

$$[n] \to (A_1, A_1^c) \to (A_1 \cap A_2, A_1 \cap A_2^c, A_1^c) \to \cdots \to (V_1(\vec{A}), \cdots, V_k(\vec{A})).$$

As an example, in the case (7.2.8) we have $\vec{A} = (A_1, A_2)$, and

$$(p \circ p) \circ p = \sum_{A_1, A_2 \subset [n]} \bar{\nu}(A_1)\bar{\nu}(A_2) p_{A_1 \cap A_2} \otimes p_{A_1 \cap A_2^c} \otimes p_{A_1^c}.$$

The proof of (7.2.7)-(7.2.9) can be easily done by induction, by splitting the tree $\gamma \in \Gamma(k)$ into the left and right subtrees γ_l, γ_r and using the relation (7.2.6), see the proof of (7.2.34) below for a closely related explicit computation.

In conclusion, from (7.2.4), (7.2.7) and (7.2.9) we obtain the following representation of the distribution p_t as a convex combination of distributions $C_\gamma^{\vec{A}}(p)$:

$$p_t = \sum_{k=1}^{\infty} \beta_t(k) \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A})\, C_\gamma^{\vec{A}}(p), \tag{7.2.13}$$

where $\beta_t(k) := e^{-t}(1 - e^{-t})^{k-1}$ is a probability on $\mathbb{N}$ for each $t \geq 0$, $\alpha_k(\gamma)$ is a probability on $\Gamma(k)$ for all k, and $\nu(\vec{A})$ is a probability on $\mathcal{V}_n^{k-1}$ for all k.

7.2.3 Contractive estimates for the nonlinear semigroup

It is convenient to use the notation $S_t(p) = p_t$ for the solution of (7.2.3) with initial datum $p_0 = p$, so that $S_{t+s} = S_t S_s = S_s S_t$, for all $s, t \geq 0$. This defines the nonlinear semigroup $\{S_t,\, t \geq 0\}$. We show a contraction property for S_t when the two initial distributions p, q have the same marginals, that is $p_i = q_i$ for all $i \in [n]$.

Theorem 7.2.2. *For any probability measure $p, q \in \mathcal{P}(\Omega)$ such that $p_i = q_i$ for all $i \in [n]$, all $t \geq 0$,*

$$\|S_t(p) - S_t(q)\|_{\mathrm{TV}} \leq \tfrac{1}{2}\, n(n-1)\, \|p - q\|_{\mathrm{TV}}\, e^{-D(\nu)t} \tag{7.2.14}$$

where $D(\nu) = 1 - r(\nu)$.

Proof. From (7.2.13) we write

$$S_t(p) - S_t(q) = \sum_{k=1}^{\infty} \beta_t(k) s^{(k)}, \tag{7.2.15}$$

where

$$s^{(k)} = \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A}) \left[C_\gamma^{\vec{A}}(p) - C_\gamma^{\vec{A}}(q) \right]. \tag{7.2.16}$$

We introduce some further notation in order to handle more explicitly the difference of probability measures $C_\gamma^{\vec{A}}(p) - C_\gamma^{\vec{A}}(q)$. For any given $\gamma \in \Gamma(k)$, and $\vec{A} \in \mathcal{V}_n^{k-1}$, $\varphi : \Omega \mapsto \mathbb{R}$, we may write

$$\begin{aligned} &\langle C_\gamma^{\vec{A}}(p) - C_\gamma^{\vec{A}}(q), \varphi \rangle \\ &= \sum_{(x,y) \in \Omega^k \times \Omega^k} \mu(x,y) \sum_{z \in \Omega} \varphi(z) \left(\mathbf{1}(z = u(y, \vec{A})) - \mathbf{1}(z = u(x, \vec{A})) \right), \end{aligned} \tag{7.2.17}$$

where $x = (x^i, i = 1, \ldots, k) \in \Omega^k$ stands for the configurations sampled with $q^{\otimes k}$ over the leaves of γ, $y = (y^i, i = 1, \ldots, k) \in \Omega^k$ stands for the configurations sampled with $p^{\otimes k}$ over the leaves of γ, and $\mu(x, y) = \prod_{i=1}^k \mu_1(x^i, y^i)$ denotes a coupling of these two product measures, such that for every $i = 1, \ldots, k$, $\mu_1(x^i, y^i)$ is the optimal coupling of (q, p):

$$\begin{aligned} \|p - q\|_{\mathrm{TV}} &= \sum_{(x,y) \in \Omega^k \times \Omega^k} \mu(x,y) \mathbf{1}(x^i \neq y^i) \\ &= \sum_{x^i, y^i \in \Omega} \mu_1(x^i, y^i) \mathbf{1}(x^i \neq y^i). \end{aligned} \tag{7.2.18}$$

The notation $u(x, \vec{A})$ in (7.2.17) is defined as follows. Note that $x^i \in \Omega$ is a vector $x^i = (x^i_1, \dots, x^i_n)$, for every i, with $x^i_\ell \in \{0, \dots, q_\ell\}$, and write $x^i_A = (x^i_\ell)_{\ell \in A}$ for the content of x^i on $A \subset [n]$. With this notation, given $x \in \Omega^k$ and $\vec{A} \in \mathcal{V}_n^{k-1}$, $u(x, \vec{A}) \in \Omega$ is defined as the unique configuration such that

$$u(x, \vec{A})_{V_i(\vec{A})} = x^i_{V_i(\vec{A})}$$

for every $i = 1, \dots, k$. In words, for each $i \in [k]$, the content of the configuration $u(x, \vec{A})$ on the set $V_i(\vec{A}) \subset V$ is taken from the configuration x^i at the i-th leaf. The validity of (7.2.17) is thus a consequence of (7.2.12).

Let $\partial\gamma$ denote the set of leaves of γ. Notice that

$$\mathbf{1}(z = u(x, \vec{A})) = \prod_{i \in \partial\gamma} \mathbf{1}(z_{V_i(\vec{A})} = x^i_{V_i(\vec{A})})$$

Let $F = F(\vec{A}, \gamma) \subset \partial\gamma$ denote the set of leaves $i \in \partial\gamma$ such that $|V_i(\vec{A})| > 1$, and write

$$\mathbf{1}(z = u(x, \vec{A})) = w_F(x, z, \vec{A}) w_{F^c}(x, z, \vec{A}) \,,$$

where for any $S \subset \partial\gamma$ we write

$$w_S(x, z, \vec{A}) = \prod_{i \in S} \mathbf{1}(z_{V_i(\vec{A})} = x^i_{V_i(\vec{A})}).$$

We write $X = (X^i)_{i=1,\dots,k}$ and $Y = (Y^i)_{i=1,\dots,k}$ for the random variables with distribution $q^{\otimes k}$ and $p^{\otimes k}$ respectively. Using the fact that μ is a product over the leaves, and the fact that p, q have the same marginals, for fixed $z \in \Omega$ we have

$$\begin{aligned} \mu[w_{F^c}(X, z, \vec{A})] &= \sum_{(x,y)} \mu(x, y) \prod_{i \in F^c} \mathbf{1}(z_{V_i(\vec{A})} = x^i_{V_i(\vec{A})}) \\ &= \sum_{(x,y)} \mu(x, y) \prod_{i \in F^c} \mathbf{1}(z_{V_i(\vec{A})} = y^i_{V_i(\vec{A})}) \\ &= \mu[w_{F^c}(Y, z, \vec{A})], \end{aligned} \tag{7.2.19}$$

and

$$\begin{aligned} \mu[w_{\partial\gamma}(Y, z, \vec{A})] &= \sum_{(x,y)} \mu(x, y) \mathbf{1}(z = u(y, \vec{A})) \\ &= \mu[w_{F^c}(X, z, \vec{A})] \mu[w_F(Y, z, \vec{A})]. \end{aligned} \tag{7.2.20}$$

With this notation we rewrite (7.2.17) as

$$\begin{aligned}\langle C_\gamma^{\vec{A}}(p) - C_\gamma^{\vec{A}}(q), \varphi\rangle &= \sum_{z\in\Omega} \varphi(z) \left[\mu[w_{\partial\gamma}(Y,z,\vec{A})] - \mu[w_{\partial\gamma}(X,z,\vec{A})\right] \\ &= \sum_{z\in\Omega} \varphi(z)\mu[w_{F^c}(X,z,\vec{A})]\, \mu\left[w_F(Y,z,\vec{A}) - w_F(X,z,\vec{A})\right].\end{aligned}$$

Note that $w_F(Y,z,\vec{A})] \neq w_F(X,z,\vec{A})$ implies that there exists $i \in F$ such that $X^i \neq Y^i$. Therefore, from (7.2.18) we see that

$$\begin{aligned}&\sum_{z\in\Omega} \mu[w_{F^c}(X,z,\vec{A})] \left|\mu\left[w_F(Y,z,\vec{A})] - w_F(X,z,\vec{A})\right]\right| \\ &\quad\le \sum_{i\in F}\sum_{z\in\Omega} \mu[w_{F^c}(X,z,\vec{A})]\, \mu\left[|w_F(Y,z,\vec{A}) - w_F(X,z,\vec{A})|\mathbf{1}(X^i \neq Y^i)\right] \\ &\quad\le 2\sum_{i\in F} \mu\left[X^i \neq Y^i\right] = 2|F|\, \|p-q\|_{\mathrm{TV}}.\end{aligned}$$

Therefore, we have shown that (7.2.16) satisfies

$$|\langle s^{(k)}, \varphi\rangle| \le 2\,\|\varphi\|_\infty\, \|p-q\|_{\mathrm{TV}} \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A})\,|F|. \tag{7.2.21}$$

To estimate (7.2.21), let us call $\mathcal{A}_j = \mathcal{A}_j(\gamma)$ the set of $\vec{A} \in \mathcal{V}_n^{k-1}$ such that $|V_j(\vec{A})| \ge 2$, and let $\mathcal{A} = \cup_{j=1}^k \mathcal{A}_j$. We have

$$\begin{aligned}\sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A})\,|F| &= \sum_{j=1}^k \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A})\,\mathbf{1}(j\in F) \\ &= \sum_{j=1}^k \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A})\,\mathbf{1}(|V_j(\vec{A})|\ge 2) \\ &= \sum_{j=1}^k \nu(\mathcal{A}_j).\end{aligned}$$

To estimate the probability $\nu(\mathcal{A}_j)$, observe that the event $\mathcal{A}_j$ implies that there exists $i_1 < i_2 \in [n]$ such that $\{i_1,i_2\} \in V_j(\vec{A})$. Moreover, recalling (7.2.11),

$$\nu\left(\{i_1,i_2\}\in V_j(\vec{A})\right) = \nu\left(\{i_1,i_2\}\in \cap_{l=1}^{d_j} A_l^j\right) = \bar{\nu}\left(\{i_1,i_2\}\in A\right)^{d_j} \le \left(\frac{r(\nu)}{2}\right)^{d_j},$$

where we use the fact that for any $i_1 < i_2$, the probability under $\bar{\nu}$ that $\{i_1,i_2\} \in A$ is bounded by $r(\nu)/2$. Indeed, conditionally on not being separated the probability that both i_1, i_2 belong to A under $\bar{\nu}$ is $1/2$ by symmetry. A union bound then

shows that

$$\nu(\mathcal{A}_j) \leq \tfrac{1}{2}\, n(n-1) \left(\frac{r(\nu)}{2}\right)^{d_j}.$$

In conclusion, if we define

$$\omega(\gamma) := \sum_{j=1}^{k} \left(\frac{r(\nu)}{2}\right)^{d_j}, \tag{7.2.22}$$

we obtain, for all $k \in \mathbb{N}$,

$$|\langle s^{(k)}, \varphi\rangle| \leq\ n(n-1)\, \|\varphi\|_\infty\, \|p-q\|_{\mathrm{TV}} \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma)\, \omega(\gamma). \tag{7.2.23}$$

From [30, Lemma 1.4], one has, for all $\varepsilon > 0$,

$$\sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma)\omega(\gamma) \leq B_\varepsilon e^{-a_\varepsilon \log k}.$$

with $a_\varepsilon = (1 - r(\nu))/(1+\varepsilon)$ and $B_\varepsilon = B_\varepsilon(r(\nu), \varepsilon)$. One could use this estimate to obtain an almost optimal decay rate a_ε. However, here we obtain the optimal exponential decay rate by computing the expectation of $\omega(\gamma)$.

Lemma 7.2.3. *For any $t \geq 0$,*

$$\sum_{k=1}^{\infty} \beta_t(k) \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma)\, \omega(\gamma) = e^{-(1-r(\nu))t}. \tag{7.2.24}$$

Assuming the validity of (7.2.24), and using $\|\mu-\mu'\|_{\mathrm{TV}} = \frac{1}{2}\max_{f:\,\|f\|_\infty\leq 1} |\mu(f)-\mu'(f)|$, from (7.2.15) and (7.2.23) we obtain

$$\|S_t(p) - S_t(q)\|_{\mathrm{TV}} \leq \tfrac{1}{2}\, n(n-1)\, \|p-q\|_{\mathrm{TV}}\, e^{-(1-r(\nu))t}, \tag{7.2.25}$$

which concludes the proof of Theorem 7.2.2.

It remains to prove (7.2.24). For any $\gamma \in \cup_{k=1}^{\infty}\Gamma(k)$, define $\mathbb{P}_t(\gamma) = \beta_t(k)\alpha_k(\gamma)$, where k is such that $\gamma \in \Gamma(k)$. Notice that for any $\gamma \in \cup_{k=1}^{\infty}\Gamma(k)$,

$$\mathbb{P}_t(\gamma) = \mathbf{1}_{\gamma=\emptyset}\, e^{-t} + \mathbf{1}_{\gamma\neq\emptyset} \int_0^t e^{-s}\mathbb{P}_{t-s}(\gamma_l)\mathbb{P}_{t-s}(\gamma_r)ds, \tag{7.2.26}$$

where $\emptyset$ denotes the empty tree with one leaf (given by the root), while for $\gamma \neq \emptyset$ we write γ_l, γ_r for the left and right subtrees. Indeed, for $\gamma = \emptyset$ (7.2.26) is imme-

diate, while for $\gamma \neq \emptyset$, supposing $\gamma \in \Gamma(k)$, and recalling (7.2.6) one has

$$\begin{aligned}\mathbb{P}_t(\gamma) &= e^{-t}(1-e^{-t})^{k-1}\alpha_k(\gamma) \\ &= e^{-t}\int_0^t e^{-(t-s)}(1-e^{-(t-s)})^{k-2}(k-1)\alpha_k(\gamma)ds \\ &= \int_0^t e^{-s}e^{-2(t-s)}(1-e^{-(t-s)})^{k-2}\alpha_j(\gamma_l)\alpha_{k-j}(\gamma_r)ds \\ &= \int_0^t e^{-s}\mathbb{P}_{t-s}(\gamma_l)\mathbb{P}_{t-s}(\gamma_r)ds.\end{aligned}$$

To prove Lemma 7.2.3, note that the left hand side of (7.2.24) is given by

$$\sum_\gamma \mathbb{P}_t(\gamma)\omega(\gamma).$$

Using (7.2.26) and $\omega(\emptyset) = 1$, we see that

$$\begin{aligned}\zeta(t) :&= \sum_\gamma \mathbb{P}_t(\gamma)\omega(\gamma) \\ &= e^{-t} + \frac{r(\nu)}{2}\sum_{\gamma\neq\emptyset}\int_0^t e^{-s}\mathbb{P}_{t-s}(\gamma_l)\mathbb{P}_{t-s}(\gamma_r)(\omega(\gamma_l)+\omega(\gamma_r))ds \\ &= e^{-t} + r(\nu)\sum_\gamma \int_0^t e^{-s}\mathbb{P}_{t-s}(\gamma)\omega(\gamma)ds \\ &= e^{-t} + r(\nu)\int_0^t e^{-s}\zeta(t-s)ds.\end{aligned}$$

Differentiating, and integrating by parts the resulting expression one finds $\dot\zeta(t) = -(1-r(\nu))\zeta(t)$. Since $\zeta(0) = 1$ we obtain $\zeta(t) = e^{-(1-r(\nu))t}$, $t \geq 0$. □

With a similar argument we obtain Theorem 6.6.2.

Proof of Theorem 6.6.2. The statement about the total variation distance is contained in Theorem 7.2.2. To prove the statement (6.6.2) about the relative entropy, we observe that by (7.2.9), (7.2.13), and convexity of the relative entropy,

$$H(p_t \,|\, \pi) \leq \sum_{k=1}^\infty \beta_t(k)\sum_{\gamma\in\Gamma(k)}\alpha_k(\gamma)\sum_{\vec A\in\mathcal{V}_n^{k-1}}\nu(\vec A)\, H\left(C_\gamma^{\vec A}(p)\,|\,\pi\right).$$

Since π is a product measure, using (7.2.12) we see that for any $\gamma \in \Gamma(k)$,

$$H\left(C_\gamma^{\vec A}(p)\,|\,\pi\right) = \sum_{j=1}^k H\left(p_{V_j(\vec A)}\,|\,\pi_{V_j(\vec A)}\right) = \sum_{j:|V_j(\vec A)|>1} H\left(p_{V_j(\vec A)}\,|\,\pi_{V_j(\vec A)}\right),$$

where use the fact that if $|V_j(\vec{A})| = 1$ then $H(p_{V_j(\vec{A})} \,|\, \pi_{V_j(\vec{A})}) = 0$ since p, π have the same marginals. Furthermore, the monotonicity property of entropy implies that

$$H\left(p_{V_j(\vec{A})} \,|\, \pi_{V_j(\vec{A})}\right) \leq H(p \,|\, \pi).$$

Therefore, repeating the argument in (7.2.23)-(7.2.24) we obtain

$$H(p_t \,|\, \pi) \leq \tfrac{1}{2}\, n(n-1)\, H(p \,|\, \pi)\, e^{-(1-r(\nu))\, t}.$$

□

Remark 7.2.4. The constant $D(\nu) = 1 - r(\nu)$ in Theorem 6.6.2 is optimal in the sense that, for any recombination measure ν, there are initial distributions p with the same marginals as π such that

$$\liminf_{t\to\infty} \frac{1}{t} \log \|p_t - \pi\|_{\mathrm{TV}} \geq -D(\nu). \tag{7.2.27}$$

To see this, pick $i_1, i_2 \in [n]$ such that $r(\nu) = \mathrm{Pr}_\nu\,(A$ does not separate i_1 and $i_2)$, see (6.1.2). For simplicity, take $\Omega = \{0,1\}^n$, π uniform over Ω and $p = \frac{1}{2}\delta_{\underline{0}} + \frac{1}{2}\delta_{\underline{1}}$. Consider the event $\{\sigma_{i_1} = \sigma_{i_2}\}$, and let $\mathcal{B}_t$ denote the event that i_1 and i_2 are not separated by the fragmentation process at time t. From (7.2.13) we write p_t as an average over the fragmentation process and note that conditionally on the event $\mathcal{B}_t$ one has $p_t(\sigma_{i_1} = \sigma_{i_2} \,|\, \mathcal{B}_t) = 1$, while conditionally on the event $\mathcal{B}_t^c$ one has $p_t(\sigma_{i_1} = \sigma_{i_2} \,|\, \mathcal{B}_t^c) = 1/2$. Moreover $\pi(\sigma_{i_1} = \sigma_{i_2}) = 1/2$ and therefore

$$\|p_t - \pi\|_{\mathrm{TV}} \geq p_t(\sigma_{i_1} = \sigma_{i_2}) - \pi(\sigma_{i_1} = \sigma_{i_2}) = \mathbb{P}(\mathcal{B}_t) + \tfrac{1}{2}\mathbb{P}(\mathcal{B}_t^c) - \tfrac{1}{2} = \tfrac{1}{2}\mathbb{P}(\mathcal{B}_t).$$

On the other hand, with the notation from (7.2.22) and (7.2.24),

$$\mathbb{P}(\mathcal{B}_t) = \sum_\gamma \mathbb{P}_t(\gamma)\omega(\gamma) = e^{-D(\nu)\, t},$$

which implies (7.2.27).

7.2.4 Contraction for the linearized equation

Consider the symmetric bilinear form $\hat{Q}(f, g)$ defined by

$$\hat{Q}(f,g)(\eta) = \frac{1}{2}\sum_{A,\sigma} \nu(A)\Big(f(\sigma_A\eta_{A^c})g(\eta_A\sigma_{A^c}) + g(\sigma_A\eta_{A^c})f(\eta_A\sigma_{A^c}) \\ - f(\sigma)g(\eta) - g(\sigma)f(\eta)\Big),$$

where $\eta \in \Omega$, $f, g : \Omega \mapsto \mathbb{R}$, and the sum extends to all $A \subset [n]$ and $\sigma \in \Omega$. If $p \in \mathcal{P}(\Omega)$, then $\hat{Q}(p,p) = Q(p)$, where

$$Q(p) = \sum_A \nu(A)(p_A \otimes p_{A^c} - p)$$

is the generator associated to (6.1.1), that is the non linear semigroup $\{S_t, t \geq 0\}$ satisfies $S_0(p) = p$ and $\partial_t S_t(p) = Q(S_t(p))$, $t \geq 0$. Next, consider the differential equation

$$\partial_t h_t = 2\hat{Q}(q_t, h_t), \qquad h_0 = h, \tag{7.2.28}$$

where $q_t = S_t(q)$ for some fixed $q \in \mathcal{P}(\Omega)$. Note that it is linear in h. We write its unique solution as

$$h_t = \bar{S}_t(q)(h),$$

that is $\bar{S}_t(q)(h)$ verifies $\partial_t \bar{S}_t(q)(h) = 2\hat{Q}(q_t, \bar{S}_t(q)(h))$, $\bar{S}_0(q)(h) = h$ and

$$\bar{S}_t(q)(\bar{S}_s(q)(h)) = \bar{S}_s(q)(\bar{S}_t(q)(h)) = \bar{S}_{t+s}(q)(h).$$

Our main result concerning the linearized equation reads as follows.

Theorem 7.2.5. *For all $p, q \in \mathcal{P}(\Omega)$ such that $p_i = q_i$ for all $i \in [n]$, and $\varphi : \Omega \mapsto \mathbb{R}$, for all $\varepsilon > 0$,*

$$\langle \bar{S}_t(q)(p-q), \varphi \rangle \leq \tfrac{1}{2}\, n(n-1) \|\varphi\|_\infty e^{-D(\nu)\, t}\, \|p - q\|_{\mathrm{TV}}, \tag{7.2.29}$$

$$\begin{aligned} &\langle S_t(p) - S_t(q) - \bar{S}_t(q)(p-q), \varphi \rangle \\ &\qquad \leq \tfrac{1}{16}\, n^3(n-1)(n-2) \|\varphi\|_\infty e^{-D(\nu)\, t}\, \|p - q\|_{\mathrm{TV}}^2 \end{aligned} \tag{7.2.30}$$

where $D(\nu) = 1 - r(\nu)$.

The proof of Theorem 7.2.5 requires several steps. We start by giving an explicit representation of the solution $\bar{S}_t(q)(p-q)$ to (7.2.28) when $h = p - q$. To this end let us define, for any $k \geq 2$, $\gamma \in \Gamma(k)$, the distribution $C_{\gamma,p}(q)$ as the measure defined in (7.2.9) with the only difference that we take the distribution p instead of q at the rightmost leaf of γ, all other leafs remaining with the distribution q. Formally,

$$C_{\gamma,p}(q) = \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A})\, C_{\gamma,p}^{\vec{A}}(q), \qquad C_{\gamma,p}^{\vec{A}}(q) = q_{V_1(\vec{A})} \otimes \cdots \otimes q_{V_{k-1}(\vec{A})} \otimes p_{V_k(\vec{A})}.$$

In what follows we denote by $d_r(\gamma)$ the depth of the rightmost leaf in γ.

Lemma 7.2.6. *For all $p, q \in \mathcal{P}(\Omega)$, $t \geq 0$,*

$$\bar{S}_t(q)(p-q) = \sum_{k=1}^{\infty} \beta_t(k) \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma)(C_{\gamma,p}(q) - C_\gamma(q))2^{d_r(\gamma)}. \tag{7.2.31}$$

Proof. Fix $p, q \in \mathcal{P}(\Omega)$. If $h_t = \bar{S}_t(q)(p)$, then

$$\partial_t \sum_\sigma h_t(\sigma) = 2 \sum_\sigma \hat{Q}(q_t, h_t)(\sigma) = 0.$$

Thus $\sum_\sigma h_t(\sigma) = 1$, and $2\hat{Q}(q_t, h_t) = 2q_t \circ h_t - q_t - h_t$. Therefore h_t satisfies

$$h_t = qe^{-t} + \int_0^t e^{-(t-s)} \left(2q_s \circ h_s - q_s\right) ds.$$

By linearity, $\bar{S}_t(q)(p-q) = \bar{S}_t(q)(p) - \bar{S}_t(q)(q)$ satisfies $\bar{S}_t(q)(p-q) = u_t - v_t$, where

$$\begin{aligned} u_t &= pe^{-t} + \int_0^t e^{-(t-s)} 2q_s \circ u_s ds, \\ v_t &= qe^{-t} + \int_0^t e^{-(t-s)} 2q_s \circ v_s ds. \end{aligned} \tag{7.2.32}$$

Wild's construction then shows that

$$u_t = e^{-t} \sum_{k=1}^{\infty} (1 - e^{-t})^{k-1} u^{(k)}, \qquad v_t = e^{-t} \sum_{k=1}^{\infty} (1 - e^{-t})^{k-1} v^{(k)}, \tag{7.2.33}$$

where $u^{(1)} = p$, $v^{(1)} = q$, and for $k \geq 2$,

$$u^{(k)} = \frac{1}{k-1} \sum_{j=1}^{k-1} 2q^{(j)} \circ u^{(k-j)}, \qquad v^{(k)} = \frac{1}{k-1} \sum_{j=1}^{k-1} 2q^{(j)} \circ v^{(k-j)}.$$

respectively, and $q^{(j)}$ is defined by (7.2.5) with p replaced by q. To check that this representation of the solution holds, let u_t be defined by (7.2.33). Then

$$\begin{aligned} \int_0^t e^{-(t-s)} 2q_s \circ u_s ds &= \sum_{\ell,k=1}^{\infty} \int_0^t \beta_s(k)\beta_s(\ell) e^{-(t-s)} 2q^{(\ell)} \circ u^{(k)} ds \\ &= \sum_{\ell<k} \int_0^t e^{-2s}(1 - e^{-s})^{k-2} e^{-(t-s)} 2q^{(\ell)} \circ u^{(k-\ell)} ds \\ &= \sum_{k \geq 2} \int_0^t (k-1) e^{-s} (1 - e^{-s})^{k-2} e^{-t} u^{(k)} ds \\ &= \sum_{k \geq 2} (1 - e^{-t})^{k-1} e^{-t} u^{(k)} = u_t - e^{-t} u^{(1)} = u_t - e^{-t} p. \end{aligned}$$

Thus u_t solves (7.2.32). The same applies to v_t. This proves (7.2.33).

Next, we prove that

$$u^{(k)} = \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) C_{\gamma,p}(q) 2^{d_r(\gamma)}\,, \qquad v^{(k)} = \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) C_{\gamma}(q) 2^{d_r(\gamma)} \quad (7.2.34)$$

for all $k \geq 2$. Let us prove this by using induction over k. If $k = 2$ then $u^{(2)} = 2q\circ p$ and thus the claimed identity holds. If the formula holds for $j \leq k-1$, then using (7.2.9) and (7.2.6),

$$\begin{aligned} u^{(k)} &= \frac{1}{k-1}\sum_{j=1}^{k-1} 2q^{(j)} \circ u^{(k-j)} \\ &= \frac{1}{k-1}\sum_{j=1}^{k-1} \sum_{\substack{\gamma_l\in\Gamma(j)\\ \gamma_r\in\Gamma(k-j)}} \alpha_j(\gamma_l)\alpha_{k-j}(\gamma_r) C_{\gamma_l}(q)\circ C_{\gamma_r,p}(q) 2^{d_r(\gamma_r)+1} \\ &= \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) C_{\gamma,p}(q) 2^{d_r(\gamma)}. \end{aligned}$$

This proves the identity (7.2.34) for u, and the same argument proves the one for v. The identities (7.2.33) and (7.2.34) imply (7.2.31). □

Next, for any $k \geq 2$, $\gamma \in \Gamma(k)$, $j \in [k]$, we define

$$C_{\gamma,p,j}(q) = \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A}) C^{\vec{A}}_{\gamma,p,j}(q)\,, \qquad C^{\vec{A}}_{\gamma,p,j}(q) := p_{V_j(\vec{A})} \otimes_{\ell\neq j} q_{V_\ell(\vec{A})},$$

which denotes the distribution obtained from the tree $\gamma \in \Gamma(k)$ when all leaves are given the distribution q except for the j-th leaf which takes the distribution p. When $j = k$, that is j is the rightmost leaf, then we have $C_{\gamma,p,k}(q) = C_{\gamma,p}(q)$.

Lemma 7.2.7. *For any $k \geq 2$, for all p, q:*

$$\sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) 2^{d_r} C_{\gamma,p}(q) = \sum_{j=1}^{k} \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) C_{\gamma,p,j}(q)\,. \quad (7.2.35)$$

Proof. By symmetry, it holds for $k = 2$, since in this case (7.2.35) says $2q \circ p =$

$q \circ p + p \circ q$. Assume that (7.2.35) holds for all $j \leq k-1$. Then

$$\begin{aligned}\sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) 2^{d_r(\gamma)} C_{\gamma,p}(q) &= \frac{2}{k-1} \sum_{j=1}^{k-1} \sum_{\gamma\in\Gamma(j)} \alpha_j(\gamma) C_\gamma(q) \\ &\qquad \circ \left(\sum_{\gamma'\in\Gamma(k-j)} \alpha_{k-j}(\gamma') 2^{d_r(\gamma')} C_{\gamma',p}(q) \right) \\ &= \frac{2}{k-1} \sum_{j=1}^{k-1} \sum_{\gamma\in\Gamma(j)} \alpha_j(\gamma) C_\gamma(q) \\ &\qquad \circ \left(\sum_{i=1}^{k-j} \sum_{\gamma'\in\Gamma(k-j)} \alpha_{k-j}(\gamma') C_{\gamma',p,i}(q) \right) \\ &= 2 \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=j(\gamma)+1}^{k} C_{\gamma,p,i}(q), \end{aligned} \tag{7.2.36}$$

where $j(\gamma)$ is the number of leaves in the left subtree of γ, and we have used (7.2.6). By symmetry,

$$\sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=j(\gamma)+1}^{k} C_{\gamma,p,i}(q) = \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=1}^{j(\gamma)} C_{\gamma,p,i}(q)\,.$$

Therefore,

$$\begin{aligned} 2 \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=j(\gamma)+1}^{k} C_{\gamma,p,i}(q) &= \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=1}^{j(\gamma)} C_{\gamma,p,i}(q) \\ &\quad + \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{i=j(\gamma)+1}^{k} C_{\gamma,p,i}(q) \\ &= \sum_{i=1}^{k} \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) C_{\gamma,p,i}(q). \end{aligned} \tag{7.2.37}$$

The desired identity follows from (7.2.36) and (7.2.37). □

We can now turn to the proof Theorem of 7.2.5.

Proof of (7.2.29). From Lemma 7.2.6 and Lemma 7.2.7, we write

$$\bar{S}_t(q)(p-q) = \sum_{k=1}^{\infty} \beta_t(k)\, \bar{s}^{(k)}, \tag{7.2.38}$$

where

$$\bar{s}^{(k)} = \sum_{j=1}^{k} \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A}) \left[C^{\vec{A}}_{\gamma,p,j}(q) - C^{\vec{A}}_{\gamma}(q) \right] .$$

Let $\mathcal{A}_j$ denote the set of $\vec{A} \in \mathcal{V}_n^{k-1}$ such that $|V_j(\vec{A})| \geq 2$. Since p, q have the same marginals, arguing as in the proof of Theorem 7.2.2, we obtain, or all $\varphi : \Omega \mapsto \mathbb{R}$,

$$|\langle C^{\vec{A}}_{\gamma,p,j}(q) - C^{\vec{A}}_{\gamma}(q), \varphi \rangle| \leq 2\|\varphi\|_\infty \|p - q\|_{\mathrm{TV}} \, \mathbf{1}(\mathcal{A}_j) \, .$$

Therefore, as in (7.2.23) we obtain

$$|\langle \bar{s}^{(k)}, \varphi \rangle| \leq n(n-1) \, \|\varphi\|_\infty \, \|p - q\|_{\mathrm{TV}} \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \omega(\gamma).$$

From (7.2.38) and Lemma 7.2.3 we conclude the proof of (7.2.29). □

Proof of (7.2.30). The proof of (7.2.30) requires a bit more work. Let us define

$$r_t := S_t(p) - S_t(q) - \bar{S}_t(q)(p - q).$$

From Lemma 7.2.6 and Lemma 7.2.7 we see that

$$r_t = \sum_{k=1}^{\infty} \beta_t(k) r^{(k)},$$

where

$$r^{(k)} = \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \left[C_\gamma(p) - C_\gamma(q) - \sum_{i=1}^{k} (C_{\gamma,p,i}(q) - C_\gamma(q)) \right] ,$$

We are going to prove that for any $\varphi : \Omega \mapsto \mathbb{R}$,

$$\langle r^{(k)}, \varphi \rangle \leq \tfrac{1}{16} \, n^3 (n-1)(n-2) \|\varphi\|_\infty \|p - q\|^2_{\mathrm{TV}} \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \omega(\gamma). \quad (7.2.39)$$

By the argument in (7.2.25) and Lemma 7.2.3 this is sufficient to end the proof.

With the notation from the proof of Theorem 7.2.2 we write

$$\langle r^{(k)}, \varphi\rangle = \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{(x,y)\in\Omega^k\times\Omega^k} \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \mu(x,y)\nu(\vec{A}) \times \tag{7.2.40}$$
$$\times \sum_{z\in\Omega} \varphi(z) \Bigg(\mathbf{1}(z = u(y,\vec{A})) - \mathbf{1}(z = u(x,\vec{A}))$$
$$- \sum_{i=1}^{k} \Big(\mathbf{1}(z = u(x,y,\vec{A},i) - \mathbf{1}(z = u(x,\vec{A})) \Big) \Bigg),$$

where we call $u(x,y,\vec{A},j)$ the unique configuration in Ω such that

$$u(x,y,\vec{A},j)_{V_j(\vec{A})} = y^j_{V_j(\vec{A})}\,, \qquad u(x,y,\vec{A},j)_{V_i(\vec{A})} = x^i_{V_i(\vec{A})}\,, \quad \text{for every } \ i\neq j.$$

That is, $u(x,y,\vec{A},j)$ coincides with $u(x,\vec{A})$ except that on $V_j(\vec{A})$ its content (if not empty) is taken from y^j. It follows that

$$\mathbf{1}(z = u(x,\vec{A})) = \prod_{i\in\partial\gamma} \mathbf{1}(z_{V_i(\vec{A})} = x^i_{V_i(\vec{A})})\,,$$
$$\mathbf{1}(z = u(x,y,\vec{A},j)) = \mathbf{1}(z_{V_j(\vec{A})} = y^j_{V_j(\vec{A})}) \prod_{i\in\partial\gamma:\ i\neq j} \mathbf{1}(z_{V_i(\vec{A})} = x^i_{V_i(\vec{A})})\,,$$

where $\partial\gamma$ denotes the set of leaves of γ. As in the proof of Theorem 7.2.2, we let $F = F(\vec{A},\gamma) \subset \partial\gamma$ denote the set of leaves $i\in\partial\gamma$ such that $|V_i(\vec{A})| > 1$, and recall that since μ is a product over the leaves, and p, q have the same marginals one has the identities (7.2.19) and (7.2.20). With this notation we rewrite (7.2.40) as

$$\langle r^{(k)}, \varphi\rangle = \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A}) \sum_{z\in\Omega} \varphi(z) \times \tag{7.2.41}$$
$$\times \Bigg[\mu[w_{\partial\gamma}(Y,z,\vec{A})] - \mu[w_{\partial\gamma}(X,z,\vec{A})]$$
$$- \sum_{i=1}^{k} \Big(\mu[w_{\partial\gamma}(X(i,y),z,\vec{A})] - \mu[w_{\partial\gamma}(X,z,\vec{A})] \Big) \Bigg],$$

where $X(i,y)\in\Omega^k$ denotes the vector whose i-th component is Y^i while all other components are X^j. Moreover, for a given choice of $\gamma, \vec{A}, z$, the square bracket in

(7.2.41) is also equal to

$$\begin{aligned}
&\mu[w_{\partial\gamma}(Y,z,\vec{A})] - \mu[w_{\partial\gamma}(X,z,\vec{A})] \qquad (7.2.42)\\
&\qquad - \sum_{i=1}^{k} \Big(\mu[w_{\partial\gamma}(X(i,y),z,\vec{A})] - \mu[w_{\partial\gamma}(X,z,\vec{A})] \Big)\\
&= \mu[w_{F^c}(X,z,\vec{A})]\ \mu\Bigg[w_F(Y,z,\vec{A}) - w_F(X,z,\vec{A})\\
&\qquad - \sum_{i\in F} \Big(w_F(X(i,y),z,\vec{A}) - w_F(X,z,\vec{A}) \Big) \Bigg].
\end{aligned}$$

In particular, if $\gamma, \vec{A}$ are such that $F = F(\gamma,\vec{A}) = \emptyset$, then (7.2.42) vanishes. Moreover, (7.2.42) vanishes also when F is a single leaf. Indeed, if e.g. $F = \{j\}$, then

$$\begin{aligned}
&w_F(Y,z,\vec{A}) - w_F(X,z,\vec{A}) - \sum_{i\in F} \Big(w_F(X(i,y),z,\vec{A}) - w_F(X,z,\vec{A}) \Big)\\
&\qquad = w_F(Y,z,\vec{A}) - w_F(X(j,y),z,\vec{A}) = 0.
\end{aligned}$$

Thus, we may restrict to the case of $|F| > 1$. Therefore,

$$|\langle r^{(k)},\varphi\rangle| \le \|\varphi\|_\infty \sum_{\gamma\in\Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A}\in\mathcal{V}_n^{k-1}} \nu(\vec{A})\mathbf{1}(|F|>1)\,\mu\Big[\Gamma(\gamma,\vec{A},X,Y)\Big],$$

where we use the notation

$$\begin{aligned}
\Gamma(\gamma,\vec{A},X,Y) = \sum_{z\in\Omega} w_{F^c}(X,z,\vec{A}) \Bigg| & w_F(Y,z,\vec{A}) - w_F(X,z,\vec{A})\\
& - \sum_{i\in F} \Big(w_F(X(i,y),z,\vec{A}) - w_F(X,z,\vec{A}) \Big) \Bigg|.
\end{aligned}$$

Clearly, if $X = Y$ the expression within absolute values above vanishes. Moreover, the same applies if (X,Y) is such that $X^i = Y^i$ for all $i \in F$ except one leaf $j \in F$ where $X^j \neq Y^j$. Indeed, in this case $X(i,y) = X$ for all $i \in F$, $i \neq j$, and $X(j,y) = Y$. Given X, Y we write $E = E(X,Y) \subset F$ for the set of leaves $i \in F$

where $X^i \neq Y^i$. Then

$$\begin{aligned}
\Gamma(\gamma, \vec{A}, X, Y) &= \sum_z w_{E^c \cup F^c}(X, z, \vec{A}) \Big| w_E(Y, z, \vec{A}) - w_E(X, z, \vec{A}) \\
&\qquad - \sum_{i \in E} \Big(w_E(X(i,y), z, \vec{A}) - w_E(X, z, \vec{A}) \Big) \Big| \\
&\leq \sum_z w_{E^c}(X, z, \vec{A}) \left| w_E(Y, z, \vec{A}) - \sum_{i \in E} w_E(X(i,y), z, \vec{A}) \right| \mathbf{1}(|E| > 1) \\
&\qquad + (|E| - 1)\mathbf{1}(|E| > 1) \\
&\leq (|E| + 1)\mathbf{1}(|E| > 1) + (|E| - 1)\mathbf{1}(|E| > 1) \\
&= 2|E|\mathbf{1}(|E| > 1) \\
&\leq 2|F|\mathbf{1}(|E| > 1).
\end{aligned}$$

Next, we use $|F| \leq n/2$, since $\cup_{i \in F} V_i(\vec{A}) \subset [n]$ and $|V_i(\vec{A})| \geq 2$ for all $i \in F$. Therefore,

$$|\langle r^{(k)}, \varphi \rangle| \leq n \|\varphi\|_\infty \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \nu(|F| > 1) \mu\left[|E| > 1\right].$$

Since μ is a product over leaves and on each leaf it satisfies (7.2.18), a union bound over the set of pairs in F shows that

$$\mu\left[|E| > 1\right] \leq \tfrac{1}{2} |F|(|F| - 1) \|p - q\|_{\mathrm{TV}}^2 \leq \tfrac{1}{8} n(n-2) \|p - q\|_{\mathrm{TV}}^2.$$

Summarizing,

$$|\langle r^{(k)}, \varphi \rangle| \leq \tfrac{1}{8} n^2(n-2) \|\varphi\|_\infty \|p - q\|_{\mathrm{TV}}^2 \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \nu(|F| > 1).$$

For a given tree γ, the argument in (7.2.23) shows that

$$\nu(|F| > 1) \leq \nu(|F| \geq 1) \leq \tfrac{1}{2} n(n-1) \sum_{i=1}^k \left(\frac{r(\nu)}{2} \right)^{d_i(\gamma)} = \tfrac{1}{2} n(n-1)\, \omega(\gamma),$$

where we use the fact that the event $F \geq 1$ coincides with $\cup_{i=1}^k \mathcal{A}_i$ from the proof of (7.2.23). This ends the proof of (7.2.39), which completes the proof of (7.2.30). $\square$

7.2.5 Monotonicity of W along the nonlinear evolution

Before proving Theorem 7.2.1, we show that the distance W appearing in that statement is monotone along the semigroup. We refer to [126] for a similar argument in the case of kinetic models. It will be convenient to work with a more symmetric version of (7.2.9). Since $p \circ q = \sum_A \bar{\nu}(A)(p \circ q)_A$ we may rewrite (7.2.9) as

$$C_\gamma(p) = \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A}) \hat{C}_\gamma^{\vec{A}}(p) \tag{7.2.43}$$

where $\nu(\vec{A})$ is defined as in (7.2.10), and $\hat{C}_\gamma^{\vec{A}}(p)$ represents the symmetric convolutions associated to the sampled sets $(A_1, \dots, A_{k-1})$. In other words, $\hat{C}_\gamma^{\vec{A}}(p)$ is defined recursively by the following relations. If $\gamma \in \Gamma(k)$ then, decomposing γ into the left and right subtrees γ_l, γ_r as in (7.2.6), one has

$$\hat{C}_\gamma^{\vec{A}}(p) = (\hat{C}_{\gamma_l}^{\vec{A}_l}(p) \circ \hat{C}_{\gamma_r}^{\vec{A}_r}(p))_{A_1}$$

where A_1 is the set attached to the root, and $\vec{A}_l := (A_2, \dots, A_j)$, $\vec{A}_r := (A_{j+1}, \dots, A_{k-1})$ are the sets associated to the left and right subtrees respectively. For example, in the case where $C_\gamma(p) = ((p \circ p) \circ p) \circ p$ as in Figure 7.2.1, one has

$$C_\gamma(p) = \sum_{A_1, A_2, A_3 \subset [n]} \bar{\nu}(A_1)\bar{\nu}(A_2)\bar{\nu}(A_3) \hat{C}_\gamma^{A_1, A_2, A_3}(p)$$

where $\hat{C}_\gamma^{A_1, A_2, A_3}(p) = (((p \circ p)_{A_3} \circ p)_{A_2}) \circ p)_{A_1}$. Therefore, as in (7.2.13) one obtains the decomposition

$$p_t = \sum_{k=1}^{\infty} \beta_t(k) \sum_{\gamma \in \Gamma(k)} \alpha_k(\gamma) \sum_{\vec{A} \in \mathcal{V}_n^{k-1}} \nu(\vec{A})\, \hat{C}_\gamma^{\vec{A}}(p). \tag{7.2.44}$$

Lemma 7.2.8. *For any $p, q \in \mathcal{P}(\Omega)$, any $k \in \mathbb{N}$, any $\gamma \in \Gamma(k)$, and any $\vec{A} \in \mathcal{V}_n^{k-1}$, one has*

$$W\left(\hat{C}_\gamma^{\vec{A}}(p), \hat{C}_\gamma^{\vec{A}}(q)\right) \leq W(p, q)\,. \tag{7.2.45}$$

In particular, $W(S_t(p), S_t(q)) \leq W(p, q)$ for any $t \geq 0$.

Proof. First we show

$$W((p_1 \circ p_2), (q_1 \circ q_2)) \leq \tfrac{1}{2} W(p_1, q_1) + \tfrac{1}{2} W(p_2, q_2) \tag{7.2.46}$$

for all $p_1, p_2, q_1, q_2 \in \mathcal{P}(\Omega)$. For a fixed $A \subset [n]$ and configurations $x_1, x_2 \in \Omega$ we

define

$$\Pi_{x_1,x_2,A}(z) := \tfrac{1}{2}\left(\mathbf{1}(z_A = x_{1,A}, z_{A^c} = x_{2,A^c}) + \mathbf{1}(z_A = x_{2,A}, z_{A^c} = x_{1,A^c})\right), \quad z \in \Omega.$$

Now we choose $\{X_1, Y_1\}, \{X_2, Y_2\}$ random variables such that

- $\mathbb{E}\left[d\left(X_1, Y_1\right)\right] = W(p_1, q_1), \mathbb{E}\left[d\left(X_2, Y_2\right)\right] = W(p_2, q_2)$.
- X_i is distributed with p_i and Y_i is distributed with $q_i, i = 1, 2$.
- $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$ are independent.

Note that

$$\mathbb{E}\left[\Pi_{X_1,X_2,A}\right] = \left(p_1 \circ p_2\right)_A.$$

Now, given $x_1, x_2, y_1, y_2 \in \Omega$ and $A \subset [n]$, we consider the following probability measure

$$\begin{aligned}\pi(z_1, z_2) :=& \tfrac{1}{2}\big(\mathbf{1}(z_{1,A} = x_{1,A}, z_{1,A^c} = x_{2,A^c}, z_{2,A} = y_{1,A}, z_{2,A^c} = y_{2,A^c}) \\ &+ \mathbf{1}(z_{1,A} = x_{2,A}, z_{1,A^c} = x_{1,A^c}, z_{2,A} = y_{2,A}, z_{2,A^c} = y_{1,A^c})\big).\end{aligned}$$

Note that this is a coupling of $\Pi_{x_1,x_2,A}$ and $\Pi_{y_1,y_2,A}$. Therefeore

$$\begin{aligned}W\left(\Pi_{x_1,x_2,A}, \Pi_{y_1,y_2,A}\right) &\leq \sum_{z_1,z_2} \pi(z_1, z_2) d(z_1, z_2) \\ &= \tfrac{1}{2}\left(d(x_{1,A}x_{2,A^c}, y_{1,A}y_{2,A^c}) + d(x_{2,A}x_{1,A^c}, y_{2,A}y_{1,A^c})\right) \\ &= \tfrac{1}{2}\left(d(x_1, y_1) + d(x_2, y_2)\right).\end{aligned}$$

Moreover, by convexity

$$W\left((p_1 \circ p_2), (q_1 \circ q_2)\right) \leq \sum_{A \subset [n]} \nu(A) W\left((p_1 \circ p_2)_A, (q_1 \circ q_2)_A\right)$$

and, for each $A \subset [n]$,

$$\begin{aligned}W\left((p_1 \circ p_2)_A, (q_1 \circ q_2)_A\right) &\leq \mathbb{E}\left[W\left(\Pi_{X_1,X_2,A}, \Pi_{Y_1,Y_2,A}\right)\right] \\ &= \tfrac{1}{2} W(p_1, q_1) + \tfrac{1}{2} W(p_2, q_2).\end{aligned} \tag{7.2.47}$$

Combining the last two inequalities we obtain (7.2.46).

We now prove (7.2.45) by induction over $k \geq 1$. Clearly, $k = 1$ is trivial. The case $k = 2$ follows by (7.2.47). Suppose that (7.2.45) holds for any $j \leq k - 1$, $\gamma \in \Gamma(j)$, $\vec{A} \in \mathcal{V}_n^{j-1}$ and let A_1 be the set attached to the root of γ. If $\gamma \in \Gamma(k)$ then

we decompose γ into the left and right subtrees γ_l, γ_r as in (7.2.6), so that

$$\hat{C}^{\vec{A}}_{\gamma}(p) = (\hat{C}^{\vec{A}_l}_{\gamma_l}(p) \circ \hat{C}^{\vec{A}_r}_{\gamma_r}(p))_{A_1}$$
$$\hat{C}^{\vec{A}}_{\gamma}(q) = (\hat{C}^{\vec{A}_l}_{\gamma_l}(q) \circ \hat{C}^{\vec{A}_r}_{\gamma_r}(q))_{A_1}.$$

where $\vec{A}_l := (A_2, \dots, A_j)$ and $\vec{A}_r := (A_{j+1}, \dots, A_{k-1})$, and j is the number of leaves of γ_l. Then, using (7.2.47) again and the inductive hypothesis we have

$$\begin{aligned} W(\hat{C}^{\vec{A}}_{\gamma}(p), \hat{C}^{\vec{A}}_{\gamma}(q)) &= W((\hat{C}^{\vec{A}_l}_{\gamma_l}(p) \circ \hat{C}^{\vec{A}_r}_{\gamma_r}(p))_{A_1}, (\hat{C}^{\vec{A}_l}_{\gamma_l}(q) \circ \hat{C}^{\vec{A}_r}_{\gamma_r}(q))_{A_1}) \\ &\leq \tfrac{1}{2} W(\hat{C}^{\vec{A}_l}_{\gamma_l}(p), \hat{C}^{\vec{A}_l}_{\gamma_l}(q)) + \tfrac{1}{2} W(\hat{C}^{\vec{A}_r}_{\gamma_r}(p), \hat{C}^{\vec{A}_r}_{\gamma_r}(q)) \leq W(p, q). \end{aligned}$$

This proves (7.2.45).

By (7.2.43) and convexity,

$$W(C_{\gamma}(p), C_{\gamma}(q)) \leq W(p, q).$$

Finally, using the expression (7.2.44) for $S_t(p) = p_t$, again by convexity we obtain

$$W(S_t(p), S_t(q)) \leq W(p, q),$$

for all $t \geq 0$. □

Remark 7.2.9. We note that (7.2.46) shows in particular that

$$W(p \circ p, q \circ q) \leq W(p, q).$$

This monotonicity is not satisfied by the total variation distance. For example let us consider $\Omega = \{0, 1\}^2$, $p := \mathbf{1}_{(1,1)}, q := \frac{1}{2}(\mathbf{1}_{(1,1)} + \mathbf{1}_{(0,0)})$ and suppose ν is the uniform crossover. Then one can check that $p \circ p = \mathbf{1}_{(1,1)}$ and $q \circ q = \frac{3}{8}(\mathbf{1}_{(1,1)} + \mathbf{1}_{(0,0)}) + \frac{1}{8}(\mathbf{1}_{(1,0)} + \mathbf{1}_{(0,1)})$, and therefore $\|p \circ p - q \circ q\|_{\mathrm{TV}} = \frac{5}{8} > \frac{1}{2} = \|p - q\|_{\mathrm{TV}}$.

7.2.6 Proof of Theorem 7.2.1

For $k \in \mathbb{N}$, the moment measures $F^{k,N}_t \in \mathcal{P}(\Omega^k)$ and $\bar{F}^{k,N}_t \in \mathcal{P}(\Omega^k)$ are defined by

$$F^{k,N}_t = \sum_{\eta \in \Omega^N} \mu_{N,t}(\eta)(\lambda_\eta)^{\otimes k}, \qquad \bar{F}^{k,N}_t = \sum_{\eta \in \Omega^N} \mu_N(\eta)(S_t\lambda_\eta)^{\otimes k}.$$

By the triangular inequality we have

$$|P_k\mu_{N,t}(\varphi_k) - p_t^{\otimes k}(\varphi_k)| \leq \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,$$

where

$$\mathcal{T}_1 := |P_k\mu_{N,t}(\varphi_k) - F_t^{k,N}(\varphi_k)|,$$
$$\mathcal{T}_2 := |F_t^{k,N}(\varphi_k) - \bar{F}_t^{k,N}(\varphi_k)|,$$
$$\mathcal{T}_3 := |\bar{F}_t^{k,N}(\varphi_k) - p_t^{\otimes k}(\varphi_k)|.$$

We are going to estimate each term separately.

The term $\mathcal{T}_1$ can be estimated by using a simple combinatorial argument, and one obtains

$$\mathcal{T}_1 \leq \frac{2k(k-1)}{N}\|\varphi_k\|_\infty.$$

Since the proof is identical to that in [37, Lemma 3.11] we omit the details.

The last term $\mathcal{T}_3$ is estimated using the initial chaos. Note that

$$p_t^{\otimes k}(\varphi_k) - \bar{F}_t^{k,N}(\varphi_k) = \sum_{\eta\in\Omega^N}\mu_N(\eta)\left(\prod_{i=1}^k\langle S_t(p),\varphi^i\rangle - \prod_{i=1}^k\langle S_t(\lambda_\eta),\varphi^i\rangle\right).$$

Using $\prod_{i=1}^k a_i - \prod_{i=1}^k b_i = \sum_{i=1}^k(\prod_{1\leq j<i}a_j)(a_i-b_i)(\prod_{i<j\leq k}b_j)$, we obtain

$$\begin{aligned}\mathcal{T}_3 \leq \sum_{\eta\in\Omega^N}\mu_N(\eta)\sum_{i=1}^k\left(\prod_{1\leq j<i}|\langle S_t(p),\varphi^j\rangle|\right)\left|\langle S_t(p)-S_t(\lambda_\eta),\varphi^i\rangle\right|\times \\ \times\left(\prod_{i<j\leq k}|\langle S_t(\lambda_\eta),\varphi^j\rangle|\right) \\ \leq 2k\|\varphi_k\|_\infty\,\mu_N\left[W\left(S_t(p),S_t(\lambda_{\chi^N})\right)\right] \leq 2k\|\varphi_k\|_\infty\,\mu_N\left[W\left(p,\lambda_{\chi^N}\right)\right],\end{aligned}$$

where χ^N has distribution μ_N, and we have used

$$|\langle\mu-\mu',f\rangle| \leq 2\|f\|_\infty\|\mu-\mu'\|_{\mathrm{TV}} \leq 2\|f\|_\infty W(\mu,\mu'),$$

for any $\mu,\mu'\in\mathcal{P}(\Omega)$, $f:\Omega\mapsto\mathbb{R}$, and the monotonicity from Lemma 7.2.8.

The estimate of the second term $\mathcal{T}_2$ is more delicate. Here we use the contraction proved in Theorem 7.2.2 and Theorem 7.2.5. Define, for $s\in[0,t]$,

$$\Psi_s^k = \sum_{\eta\in\Omega^N}\mu_{N,t-s}(\eta)(S_s(\lambda_\eta))^{\otimes k}.$$

Then $\Psi_s^k\in\mathcal{P}(\Omega^k)$ for all $s\in[0,t]$, and

$$F_t^{k,N}(\varphi_k) = \Psi_0^k(\varphi_k), \qquad \bar{F}_t^{k,N}(\varphi_k) = \Psi_t^k(\varphi_k).$$

Therefore,

$$\bar{F}_t^{k,N}(\varphi_k) - F_t^{k,N}(\varphi_k) = \int_0^t \partial_s \left[\Psi_s^k(\varphi_k)\right] \, ds \, .$$

Now,

$$\begin{aligned}
\partial_s \left[\Psi_s^k(\varphi_k)\right] &= \sum_{\eta\in\Omega^N} (\partial_s \mu_{N,t-s}(\eta))(S_s(\lambda_\eta))^{\otimes k}(\varphi_k) \\
&\quad + \sum_{\eta\in\Omega^N} \mu_{N,t-s}(\eta)\partial_s (S_s(\lambda_\eta))^{\otimes k}(\varphi_k) \\
&= -\sum_{\eta\in\Omega^N} \mu_{N,t-s}(\eta)\mathcal{L}_N \prod_{\ell=1}^k \langle S_s(\lambda_\eta), \varphi^\ell\rangle \\
&\quad + \sum_{\eta\in\Omega^N} \mu_{N,t-s}(\eta) \sum_{i=1}^k \langle Q(S_s(\lambda_\eta)), \varphi^i\rangle \prod_{\ell\neq i} \langle S_s(\lambda_\eta, \varphi^\ell\rangle.
\end{aligned} \tag{7.2.48}$$

Next, we show that

$$Q(S_s(\lambda_\eta)) = \frac{1}{N}\sum_{i<j}\sum_A \nu(A)\bar{S}_s(\lambda_\eta)(\lambda_{\eta^{i,j,A}} - \lambda_\eta). \tag{7.2.49}$$

Consider the linearized equation (7.2.28). Taking $q \in \mathcal{P}(\Omega)$ and $h = Q(q)$ one has that the solution $h_t = \bar{S}_t(q)(h)$ satisfies $h_t = Q\left(S_t(q)\right)$, that is

$$\bar{S}_t(q)(Q(q)) = Q\left(S_t(q)\right) \, , \qquad t \geq 0, \tag{7.2.50}$$

for all $q \in \mathcal{P}(\Omega)$. To see this note that for all $t \geq 0$,

$$\sum_{\sigma\in\Omega} S_t(q)(\sigma) = 1 \, , \qquad \sum_{\sigma\in\Omega} Q(S_t(q))(\sigma) = 0 \, ,$$

and therefore

$$\begin{aligned}
\partial_t Q\left(S_t(q)\right)(\sigma') &= \sum_{\substack{A\subset[n]\\ \sigma\in\Omega}} \nu(A)\Big(Q(S_t(q))(\sigma_{A^c}\sigma'_A)S_t(q)(\sigma_A\sigma'_{A^c}) \\
&\qquad + S_t(q)(\sigma_{A^c}\sigma'_A)Q(S_t(q))(\sigma_A\sigma'_{A^c}) - Q(S_t(q))(\sigma')S_t(q)(\sigma)\Big) \\
&= \sum_{\substack{A\subset[n]\\ \sigma\in\Omega}} \nu(A)\Big(Q(S_t(q))(\sigma_{A^c}\sigma'_A)S_t(q)(\sigma_A\sigma'_{A^c}) + S_t(q))(\sigma_{A^c}\sigma'_A)Q(S_t(q))(\sigma_A\sigma'_{A^c}) \\
&\qquad - Q(S_t(q))(\sigma')S_t(q)(\sigma) - Q(S_t(q))(\sigma)S_t(q)(\sigma')\Big) \\
&= 2\hat{Q}(S_t(q), Q(S_t(q)))(\sigma').
\end{aligned}$$

This proves (7.2.50). If $q = \lambda_\eta$ for some $\eta \in \Omega^N$, then

$$Q(\lambda_\eta) = \sum_A \nu(A) \left((\lambda_\eta)_A \otimes (\lambda_\eta)_{A^c} - \lambda_\eta \right).$$

Moreover, for any $A \subset [n]$,

$$\begin{aligned}(\lambda_\eta)_A \otimes (\lambda_\eta)_{A^c} - \lambda_\eta &= \frac{1}{2N^2} \sum_{i,j=1}^N (\mathbf{1}_{\eta^{i,j,A}(i)} + \mathbf{1}_{\eta^{i,j,A}(j)} - \mathbf{1}_{\eta(i)} - \mathbf{1}_{\eta(j)}) \\ &= \frac{1}{N} \sum_{i<j} (\lambda_{\eta^{i,j,A}} - \lambda_\eta).\end{aligned}$$

It follows that

$$Q(\lambda_\eta) = \frac{1}{N} \sum_{i<j} \sum_A \nu(A) (\lambda_{\eta^{i,j,A}} - \lambda_\eta).$$

Thus, using (7.2.50) and the linearity of $\bar{S}_s(\lambda_\eta)(\cdot)$ we obtain (7.2.49). On the other hand,

$$\mathcal{L}_N \prod_{\ell=1}^k \langle S_s(\lambda_\eta), \varphi^\ell \rangle = \frac{1}{N} \sum_{i<j} \sum_A \nu(A) \left(\prod_{\ell=1}^k \langle S_s(\lambda_{\eta^{i,j,A}}), \varphi^\ell \rangle - \prod_{\ell=1}^k \langle S_s(\lambda_\eta), \varphi^\ell \rangle \right).$$

Suppose we can show that

$$\begin{aligned}\Big| &\prod_{\ell=1}^k \langle S_s(\lambda_{\eta^{i,j,A}}), \varphi^\ell \rangle - \prod_{\ell=1}^k \langle S_s(\lambda_\eta), \varphi^\ell \rangle \\ &- \sum_{u=1}^k \langle \bar{S}_s(\lambda_\eta)(\lambda_{\eta^{i,j,A}} - \lambda_\eta), \varphi^u \rangle \prod_{\ell \neq u} \langle S_s(\lambda_\eta, \varphi^\ell \rangle \Big| \leq K_s, \end{aligned} \tag{7.2.51}$$

for some function K_s, $s \in [0, \infty)$, independent of η, i, j, A . Then by (7.2.48) and (7.2.49) we would have $\mathcal{T}_2 \leq \frac{N}{2} \int_0^t K_s ds$. Thus the proof will be completed by proving (7.2.51) for a suitable K_s.

Observe that

$$\begin{aligned}\prod_{\ell=1}^{k} a_\ell - \prod_{\ell=1}^{k} b_\ell - \sum_{u=1}^{k} c_u \prod_{j\neq u} b_j &= \sum_{u=1}^{k} \left[(\prod_{1\le j<u} a_j)(a_u - b_u) - (\prod_{1\le j<u} b_j)c_u \right] \prod_{j>u} b_j \\ &= \sum_{u=1}^{k} (a_u - b_u - c_u) \prod_{j\neq u} b_j + \sum_{u=1}^{k} \left[(\prod_{1\le j<u} a_j) - (\prod_{1\le j<u} b_j) \right] (a_u - b_u) \prod_{j>u} b_j \\ &= \sum_{u=1}^{k} (a_u - b_u - c_u) \prod_{j\neq u} b_j + \sum_{u=1}^{k} \sum_{v=1}^{u-1} (\prod_{j<v} a_j)(\prod_{v<j<u} b_j)(a_v - b_v)(a_u - b_u) \prod_{j>u} b_j.\end{aligned}$$

Define $a_\ell = \langle S_s(\lambda_{\eta^{i,j,A}}), \varphi^\ell \rangle$, $b_\ell = \langle S_s(\lambda_\eta), \varphi^\ell \rangle$, and $c_\ell = \langle \bar{S}_s(\lambda_\eta)(\lambda_{\eta^{i,j,A}} - \lambda_\eta), \varphi^\ell \rangle$. Noticing that

$$\|\lambda_{\eta^{i,j,A}} - \lambda_\eta\|_{\mathrm{TV}} \le \frac{2}{N},$$

and that $\lambda_{\eta^{i,j,A}}, \lambda_\eta$ have the same marginals, it follows from Theorem 7.2.5 that

$$|a_u - b_u - c_u| \le n^5 \|\varphi^u\|_\infty e^{-D(\nu)\, s} \frac{4}{N^2}.$$

Moreover, Theorem 7.2.2 shows that for all u, v,

$$|a_v - b_v||a_u - b_u| \le n^4 \|\varphi^v\|_\infty \|\varphi^u\|_\infty e^{-2D(\nu)\, s} \frac{4}{N^2}.$$

These bounds hold uniformly in $\eta \in \Omega^N, 1 \le i < j \le n, A \subset [n]$. This proves (7.2.51) with

$$K_s = k e^{-D(\nu)\, s} (n^5 + (k-1) n^4 e^{-D(\nu)\, s}) \frac{\|\varphi_k\|_\infty}{N^2}.$$

Therefore,

$$\mathcal{T}_2 \le \frac{N}{2} \int_0^t K_s \le \frac{k^2 n^5}{D(\nu) N} \|\varphi_k\|_\infty (1 - e^{-D(\nu)\, t}).$$

The proof of Theorem 7.2.1 is complete.

7.3 Relative entropy decay

The goal of this section is to prove Theorem 6.5.2, Proposition 6.5.5, and Theorem 6.6.1. To simplify our notation we write $\mu = \pi_N$ for the uniform distribution over

$\mathcal{S}_{N,n}$. For any $A \subset [n]$, $f : \mathcal{S}_{N,n} \mapsto \mathbb{R}$, we are going to use the notation

$$\mu_A f = \mu(f \,|\, \eta_{A^c}),$$

where $\mu(\cdot \,|\, \eta_{A^c})$ denotes the conditional expectation given the variables

$$\eta_{A^c} := \{\eta_i(j),\ i \in A^c,\ j = 1, \ldots, N\}.$$

The function $\mu_A f$ thus depends on η only through the variabes η_{A^c}. When $A = [n]$ we have the global expectation $\mu_{[n]} f = \mu(f)$. With this notation, μ_A is the orthogonal projection, in $L^2(\mathcal{S}_{N,n}, \mu)$, onto the space of functions that do not depend on the A-component of $\eta \in \mathcal{S}_{N,n}$. Notice the relations $\mu(f) = \mu(\mu_A f)$, and $\mu_B(f) = \mu_B(\mu_A f)$, for all $A \subset B \subset [n]$. We also use the notation $\mathrm{Ent}_A(f)$ for the entropy of $f : \mathcal{S}_{N,n} \mapsto \mathbb{R}_+$ with respect to μ_A, that is

$$\mathrm{Ent}_A(f) = \mu_A \left[f \log(f/\mu_A f) \right] .$$

Thus $\mathrm{Ent}_A(f)$ is a function that depends on η through the variables η_{A^c} only. Its expectation satisfies

$$\mu \left[\mathrm{Ent}_A(f) \right] = \mathrm{Ent}(f) - \mathrm{Ent}[\mu_A f] \,,$$

where $\mathrm{Ent}(f) = \mathrm{Ent}_{[n]}(f) = \mu \left[f \log(f/\mu(f)) \right]$.

We recall the following well known tensorization property, which is an immediate consequence of Lemma 2.4.4:

$$\mathrm{Ent}(f) \leq \sum_{i=1}^{n} \mu \left[\mathrm{Ent}_i(f) \right] . \tag{7.3.1}$$

7.3.1 The case $n = 1$

A key ingredient of our proof is the control of the base case $n = 1$. Here the problem reduces to standard random transpositions and one can use a well known bound that was first derived in [66, 64]. In our setting it can be summarized as follows, see [64, Theorem 1] or [66, Corollary 3.1] for a proof. Note that for $n = 1$ we have $\mathcal{S}_{N,n} = S_N$.

Lemma 7.3.1. *for all $N \geq 2$, for all $g : S_N \mapsto \mathbb{R}_+$, for any $i = 1, \ldots, n$,*

$$\sum_{\tau \in S_N} g(\tau) \log(g(\tau)/\bar{g}) \leq \frac{1}{N} \sum_{j<\ell} \sum_{\tau \in S_N} \big(g(\tau^{j,\ell}) - g(\tau) \big) \log \frac{g(\tau^{j,\ell})}{g(\tau)},$$

where $\bar{g} = \frac{1}{N!} \sum_{\tau \in S_N} g(\tau)$, and $\tau^{j,\ell}$ denotes τ composed with the transposition at $\{j, \ell\}$.

The proof of Theorem 6.5.2 is based on Lemma 2.4.4, Lemma 7.3.1 and the following analysis of the entropy associated to partial random permutations of the particle configuration, which is the main novelty in this section.

7.3.2 Permutation entropies

For any $A \subset [n]$, $g : \mathcal{S}_{N,n} \mapsto \mathbb{R}$, define $P_A g : \mathcal{S}_{N,n} \mapsto \mathbb{R}$, as

$$P_A g(\eta) = \frac{1}{N!} \sum_{\tau \in S_N} g((\tau\eta)_A \, \eta_{A^c}),$$

where $\tau\eta := \tau \circ \eta$ denotes the element of $\mathcal{S}_{N,n}$ obtained from η by permuting the particle labels according to τ:

$$(\tau\eta)(j) = \eta(\tau(j)).$$

The linear operator P_A is the orthogonal projection, in $L^2(\mathcal{S}_{N,n}, \mu)$, onto the space of functions that are symmetric w.r.t. permutations restricted to the subset A. When $A = \{i\}$ we write $P_{\{i\}} = P_i$, and note that $P_i g = \mu(g \,|\, \eta_{\{i\}^c}) = \mu_i g$ for all i. Note also that P_A, P_B do not commute for general $A, B \subset [n]$, but if $A \cap B = \emptyset$ then $P_A P_B = P_B P_A$. Moreover, one easily checks that if $A \subset B \subset [n]$, then

$$P_A P_B = P_{B\setminus A} P_A = P_A P_{B\setminus A} = P_B P_A, \qquad A \subset B. \tag{7.3.2}$$

Also, observe that the orthogonal projection μ_A defined above satisfies

$$\mu_A = \prod_{i \in A} P_i.$$

Notice that $P_{[n]} g = g$ iff $g \in \mathbb{S}$, and that in this case

$$P_A g = P_{A^c} g = P_A P_{A^c} g = P_{A^c} P_A g\,, \qquad g \in \mathbb{S}. \tag{7.3.3}$$

For any fixed $f \geq 0$, and $A \subset [n]$, define

$$\varphi(A; f) = \mu\left[f \log(f / P_A f)\right].$$

The function φ is a suitable conditional entropy of f on A. Namely, $\varphi(A; f)$ is the expected value with respect to μ of the entropy

$$P_A\left[f \log(f / P_A f)\right].$$

In particular, $\varphi(A; f) \geq 0$. We call $\varphi(A; f)$ the *permutation entropy* of f on A. When there is no risk of confusion we write simply $\varphi(A)$ for $\varphi(A; f)$. In general, one has

Lemma 7.3.2. *Fix a function $f \geq 0$. For any $i \in [n]$, $\varphi(\{i\}) = \mu[\mathrm{Ent}_i(f)]$, and for all $A \subset [n]$:*

$$\varphi(A) = \mu[\mathrm{Ent}_A(f)] - \mu[\mathrm{Ent}_A(P_A f)] \ . \tag{7.3.4}$$

Moreover, for all $A \subset [n]$

$$\varphi(A) \leq 2\mathcal{E}_A(f, \log f), \tag{7.3.5}$$

where

$$\mathcal{E}_A(f, \log f) = \frac{1}{2}\frac{1}{N}\sum_{j<\ell} \mu\left[\left(f^{j,\ell,A} - f\right)\log\frac{f^{j,\ell,A}}{f}\right].$$

Proof. If $A = \{i\}$, then $P_A f = \mu_i f$ and therefore $\varphi(\{i\}) = \mu[\mathrm{Ent}_i(f)]$. In general, for any A,

$$\begin{aligned}\mu[\mathrm{Ent}_A(f)] &= \mu[f\log(f/\mu_A f)] \\ &= \mu[f\log(f/P_A f)] + \mu[f\log(P_A f/\mu_A f)]\,.\end{aligned}$$

The second term above satisfies

$$\mu[f\log(P_A f/\mu_A f)] = \mu[P_A f\log(P_A f/\mu_A P_A f)] = \mu[\mathrm{Ent}_A(P_A f)]\ ,$$

where we have used that $\mu_B P_A = \mu_B$ for any $A \subset B \subset [n]$ since P_A is an orthogonal projection. This proves (7.3.4).

To prove (7.3.5), observe that for any $A \subset [n]$, $f \geq 0$, any fixed $\eta \in \mathcal{S}_{N,n}$, taking $g(\tau) := f((\tau\circ\eta)_A\eta_{A^c})$, one has $\bar{g} = P_A f$, and therefore by Lemma 7.3.1,

$$\begin{aligned}&P_A[f\log(f/P_A f)](\eta) = \frac{1}{N!}\sum_{\tau\in S_N} g(\tau)\log(g(\tau)/\bar{g}) \\ &\leq \frac{1}{N}\sum_{j<\ell}\frac{1}{N!}\sum_{\tau\in S_N}\left(f((\tau^{j,\ell}\eta)_A\eta_{A^c}) - f((\tau\eta)_A\eta_{A^c})\right)\log\frac{f((\tau^{j,\ell}\eta)_A\eta_{A^c})}{f((\tau\eta)_A\eta_{A^c})} \\ &= \frac{1}{N}\sum_{j<\ell} P_A\left[\left(f^{j,\ell,A} - f\right)\log\frac{f^{j,\ell,A}}{f}\right](\eta).\end{aligned}$$

Taking the expectation and using $\mu P_A = \mu$ one obtains (7.3.5). □

Next, we compare the permutation entropy $\varphi(A)$ with $\mu[\mathrm{Ent}_A(f)]$. The previous lemma in particular shows that $\varphi(A) \leq \mu[\mathrm{Ent}_A(f)]$. The next lemma allows us to give a bound in the opposite direction. Notice that such a bound

needs some care since if e.g. $f \in \mathbb{S}$ is not a constant then $\varphi([n]; f) = 0$ while $\mu\left[\mathrm{Ent}_{[n]}(f)\right] = \mathrm{Ent}(f) \neq 0$.

Lemma 7.3.3. *Fix $A \subset [n]$ such that $0 \leq |A| \leq n-1$, and $V \subset A^c$. Then,*

$$\varphi(A) + \sum_{i \in V} \varphi(A \cup \{i\}) \geq \mu\left[\mathrm{Ent}_V f\right].$$

Proof. Write $V = \{x_1, \ldots, x_k\} \subset [n]$, $k = |V|$, and define $A_0 = A$, and $A_i = A \cup \{x_i\}$, $i = 1, \ldots, k$. In general P_{A_i} and P_{A_j} do not commute for $i, j = 1, \ldots, k$, but using (7.3.2) one has $P_{A_0} P_{A_i} = P_{A_0} P_{x_i} = P_A \mu_{x_i}$ and for any $\ell = 1, \ldots, k$,

$$P^\ell := P_{A_0} P_{A_1} \cdots P_{A_\ell} = P_A \prod_{i=1}^{\ell} \mu_{x_i} = \left(\prod_{i=1}^{\ell} \mu_{x_i} \right) P_A = \prod_{i=0}^{\ell} P_{A_i},$$

where the last identity holds regardless of the order of the multiplications. In other words, the operators P_{A_i} commute thanks to the presence of P_{A_0}. In particular, $P^k = P_A \mu_V = \mu_V P_A$.

Consider the entropy

$$\mu\left[f \log \frac{f}{P^k f} \right] = \mu\left[f \log \frac{f}{\mu_V f} \right] + \mu\left[f \log \frac{\mu_V f}{P_A \mu_V f} \right]. \tag{7.3.6}$$

The first term is $\mu\left[\mathrm{Ent}_V f\right]$. The second term satisfies

$$\mu\left[f \log \frac{\mu_V f}{P_A \mu_V f} \right] = \mu\left[\mu_V f \log \frac{\mu_V f}{\mu_V P_A f} \right] = \varphi(A; \mu_V f) \geq 0.$$

On the other hand,

$$\log \frac{f}{P^k f} = \log \frac{f}{P_{A_0} f} + \log \frac{P_{A_0} f}{P_{A_0} P_{A_1} f} + \cdots + \log \frac{P_{A_0} P_{A_1} \cdots P_{A_{k-1}} f}{P_{A_0} P_{A_1} \cdots P_{A_k} f},$$

and therefore

$$\mu\left[f \log \frac{f}{P^k f} \right] = \varphi(A; f) + \sum_{\ell=0}^{k-1} \mu\left[f \log \frac{P^\ell f}{P^{\ell+1} f} \right]. \tag{7.3.7}$$

Next, we show that

$$\mu\left[f \log \frac{P^\ell f}{P^{\ell+1} f} \right] \leq \varphi\left(A_{\ell+1}; f\right), \tag{7.3.8}$$

for all $\ell = 0, \ldots, k-1$. Combined with (7.3.6)-(7.3.7), this implies the desired conclusion:

$$\varphi(A) + \sum_{i \in V} \varphi(A \cup \{i\}) \geq \mu\left[\mathrm{Ent}_V f\right] + \varphi(A; \mu_V f) \geq \mu\left[\mathrm{Ent}_V f\right].$$

It remains to prove (7.3.8). We first observe that

$$\mu\left[f\log\frac{P^\ell f}{P^{\ell+1}f}\right]=\mu\left[P^\ell f\log\frac{P^\ell f}{P^{\ell+1}f}\right]=\varphi\left(A_{\ell+1};P^\ell f\right). \tag{7.3.9}$$

The well known variational principle for the relative entropy implies that for any $B\subset[n]$

$$P_B\left(f\log\frac{f}{P_Bf}\right)\geq P_B(fg),$$

for any function g such that $P_B(e^g)=1$. Choosing $g=\log\frac{P^\ell f}{P_BP^\ell f}$ shows that

$$P_B\left(f\log\frac{f}{P_Bf}\right)\geq P_B\left(f\log\frac{P^\ell f}{P_BP^\ell f}\right).$$

Taking the expectation one finds

$$\varphi(B;f)\geq\mu\left[f\log\frac{P^\ell f}{P_BP^\ell f}\right].$$

If P_B,P^ℓ commute, then

$$\mu\left[f\log\frac{P^\ell f}{P_BP^\ell f}\right]=\mu\left[f\log\frac{P^\ell f}{P^\ell P_Bf}\right]=\mu\left[P^\ell f\log\frac{P^\ell f}{P^\ell P_Bf}\right]=\varphi(B;P^\ell f).$$

Therefore,

$$\varphi(B;f)\geq\varphi(B;P^\ell f),$$

whenever P_B,P^ℓ commute. Taking $B=A_{\ell+1}$, and using the fact that $P_{A_{\ell+1}}$ and P^ℓ commute, one obtains $\varphi\left(A_{\ell+1};P^\ell f\right)\leq\varphi\left(A_{\ell+1};f\right)$. Together with (7.3.9), this implies (7.3.8). □

7.3.3 Proof of Theorem 6.5.2

From the strict separation assumption, it follows that for some $\delta(\nu)>0$,

$$\mathcal{E}_{N,n}(f,\log f)\geq\frac{\delta(\nu)}{n}\sum_{i=1}^n\left(\mathcal{E}_{A_i}(f,\log f)+\mathcal{E}_{A_i\setminus\{i\}}(f,\log f)\right),$$

where, for every i, $A_i\subset[n]$ is such that $A_i\ni i$ and $\min\{\nu(A_i),\nu(A_i\setminus\{i\})\}\geq\delta(\nu)$. Therefore, from Lemma 7.3.2,

$$\mathcal{E}_{N,n}(f,\log f)\geq\frac{\delta(\nu)}{2n}\sum_{i=1}^n\left(\varphi(A_i)+\varphi(A_i\setminus\{i\})\right).$$

Lemma 7.3.3 then implies

$$\mathcal{E}_{N,n}(f, \log f) \geq \frac{\delta(\nu)}{2n} \sum_{i=1}^{n} \mu\left[\mathrm{Ent}_i f\right].$$

Using (7.3.1) it follows that $\alpha(N,n) \geq \delta(\nu)/2n$. This proves the bound (6.5.2) with $\alpha(\nu) = \delta(\nu)/2n$.

To prove the lower bound for one-point crossover, notice that the above argument can be repeated but this time we can directly estimate

$$\mathcal{E}_{N,n}(f, \log f) \geq \frac{1}{4}\frac{1}{n+1} \sum_{i=1}^{n} (\varphi(J_i) + \varphi(J_{i-1})) \geq \frac{1}{4(n+1)} \sum_{i=1}^{n} \mu\left[\mathrm{Ent}_i f\right],$$

where $J_0 = \emptyset$, and $J_i = \{1, \ldots, i\}$, $i \geq 1$. The lower bound $\alpha(N,n) \geq 1/4(n+1)$ thus follows again by the tensorization (7.3.1).

Next, we prove the lower bound for the case of uniform crossover $\nu(A) = 2^{-n}$ for all $A \subset [n]$. From Lemma 7.3.2

$$\mathcal{E}_{N,n}(f, \log f) = 2^{-n} \sum_{A} \mathcal{E}_A(f, \log f) \geq 2^{-n-1} \sum_{A} \varphi(A).$$

By Lemma 7.3.3,

$$\begin{aligned}\sum_{i=1}^{n} \sum_{A} \varphi(A) \mathbf{1}(i \in A) &= \sum_{i=1}^{n} \sum_{A\colon |A| \leq n-1} \varphi(A \cup \{i\}) \mathbf{1}(i \notin A) \\ &\geq \sum_{A\colon |A| \leq n-1} \left(\mu\left[\mathrm{Ent}_{A^c} f\right] - \varphi(A)\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{A} \varphi(A) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{A} \varphi(A)(\mathbf{1}(i \notin A) + \mathbf{1}(i \in A)) \\ &\geq \frac{1}{n} \sum_{A} |A^c| \varphi(A) + \frac{1}{n} \sum_{A\colon |A| \leq n-1} \left(\mu\left[\mathrm{Ent}_{A^c} f\right] - \varphi(A)\right) \\ &= \frac{1}{n} \sum_{A\colon |A| \leq n-1} (|A^c| - 1)\varphi(A) + \frac{1}{n} \sum_{A} \mu\left[\mathrm{Ent}_{A^c} f\right].\end{aligned}$$

In particular,

$$2^{-n-1} \sum_{A} \varphi(A) \geq \frac{2^{-n}}{2n} \sum_{A} \mu\left[\mathrm{Ent}_{A^c} f\right].$$

From Lemma 2.4.4 it follows that

$$2^{-n-1}\sum_A \varphi(A) \geq \frac{1}{4n}\mathrm{Ent} f.$$

This proves the desired bound $\alpha(N,n) \geq \frac{1}{4n}$.

Finally, in the case of symmetric functions one has $\varphi(A) = \varphi(A^c)$ for all $A \subset [n]$, see (7.3.3). Therefore, the previous computation now shows that

$$\sum_A \varphi(A) \geq \frac{2}{n}\sum_A \mu\left[\mathrm{Ent}_{A^c} f\right] - \frac{2}{n}\sum_A \varphi(A).$$

Rearranging terms yields the bound $\alpha_{\mathbb{S}}(N,n) \geq \frac{1}{2(n+2)}$. This concludes the proof of Theorem 6.5.2.

7.3.4 Proof of Proposition 6.5.5

The proof of the upper bound on $\alpha(N,n)$ is based on Proposition 7.1.11, Proposition 7.1.8, and the following entropy production estimate for the nonlinear equation that was derived in [29].

Lemma 7.3.4. *Let $\Omega = \{0,1\}^n$, and let $p = p(n) \in \mathcal{P}(\Omega)$ be defined as*

$$p = w^2\delta_{\underline{1}} + (1-w)^2\delta_{\underline{0}} + 2w(1-w)\,\mathcal{U},$$

where $w = 2^{-n}$ and $\mathcal{U}$ is the uniform distribution on Ω. Then, taking $f = p/\pi$, with $\pi = \otimes_{i=1}^n p_i$, one has

$$\frac{D_\pi(f)}{\mathrm{Ent}_\pi f} \leq \frac{4}{n} + O\left(\frac{1}{n^2}\right).$$

Proof. See [29, Proposition 4.7]. □

To prove Proposition 6.5.5 we take $\Omega = \{0,1\}^n$, and p as in Lemma 7.3.4. Note that p has marginals equal to Bernoulli $\mathrm{Be}(w)$. Take ϱ_N such that (7.1.3) holds with π the product of Bernoulli $\mathrm{Be}(w)$, and write $f_N = \gamma(p,\varrho_N)/\gamma(\pi,\varrho_N)$. Recall that

$$\alpha(N,n) \leq \alpha(\Omega_{\varrho_N}) \leq \frac{\mathcal{E}(f_N, \log f_N)}{\mathrm{Ent} f_N},$$

where $\mathrm{Ent} f_N = H_N(\gamma(p,\varrho_N)\,|\,\gamma(\pi,\varrho_N))$. Clearly, p is irreducible. From Proposition 7.1.11 and Proposition 7.1.8,

$$\limsup_{N\to\infty}\ \alpha(N,n) \leq \frac{D_\pi(f)}{\mathrm{Ent}_\pi f} \leq \frac{4}{n} + O\left(\frac{1}{n^2}\right).$$

7.3.5 Proof of Theorem 6.6.1

Let $p \in \mathcal{P}(\Omega)$ be an arbitrary initial value for the Boltzmann equation and let $\pi = \otimes_i p_i$ denote the corresponding equilibrium. In order to ensure that p be irreducible we write

$$p^{(\varepsilon)} = \varepsilon\, \pi + (1-\varepsilon)\, p$$

with $\varepsilon \in (0,1)$ to be taken to zero eventually. Clearly, $p, p^{(\varepsilon)}, \pi$ have the same marginals. Let ϱ_N be an admissible sequence such that (7.1.3) holds. Let $p_N^{(\varepsilon)} = \gamma(p^{(\varepsilon)}, \varrho_N)$ and $\pi_N = \gamma(\pi, \varrho_N)$ and define $p_{N,t}^{(\varepsilon)} = p_N^{(\varepsilon)} e^{t\mathcal{L}_N}$. The propagation of chaos at fixed time t implies that the hypothesis of Proposition 7.1.9 apply to $\mu^{(N)} = p_{N,t}^{(\varepsilon)}$. Therefore,

$$H(p_t^{(\varepsilon)} \,|\, \pi) \leq \liminf_{N\to\infty} \frac{H_N(p_{N,t}^{(\varepsilon)} \,|\, \pi_N)}{N},$$

where $p_t^{(\varepsilon)}$ is the solution to the nonlinear equation with initial datum $p^{(\varepsilon)}$. Theorem 6.5.2 implies that

$$H_N(p_{N,t}^{(\varepsilon)} \,|\, \pi_N) \leq e^{-\alpha(\nu)\, t} H_N(p_N^{(\varepsilon)} \,|\, \pi_N).$$

Then an application of Proposition 7.1.8 shows that for all $\varepsilon > 0$, $t \geq 0$, one has

$$H(p_t^{(\varepsilon)} \,|\, \pi) \leq e^{-\alpha(\nu)\, t} H(p^{(\varepsilon)} \,|\, \pi).$$

The conclusion follows by taking $\varepsilon \to 0^+$. Indeed, $p^{(\varepsilon)} \to p$ and, by (7.2.14) we know that $p_t^{(\varepsilon)} \to p_t$, so that $H(p^{(\varepsilon)} \,|\, \pi) \to H(p \,|\, \pi)$, and $H(p_t^{(\varepsilon)} \,|\, \pi) \to H(p_t \,|\, \pi)$. We note that we can use the entropy production for symmetric functions here, so that for instance in the case of uniform crossover we may take $\alpha(\nu) \geq 1/2(n+2)$.

Appendix A

Appendix

Here we prove some facts stated in chapter 2 and the local central limit theorem in chapter 7.

A.1 Mixing time and entropy

In order to prove Lemma 2.3.3, we need the following result which shows that the bound (2.3.5) is stronger than the MLSI in (2.3.3).

Lemma A.1.1. *If the entropy decay holds with rate δ in discrete time then it holds with the same rate in continuous time. That is,* (2.3.5) *implies the MLSI with constant δ.*

proof of Lemma A.1.1. Suppose that

$$\mathrm{Ent}_\mu(P^* f) \leq (1-\delta)\mathrm{Ent}_\mu f.$$

From the variational principle (2.3.2) it follows that for any $f \geq 0$ with $\mu[f] = 1$:

$$\mu[(P^* f)\log f] \leq \mathrm{Ent}_\mu P^* f.$$

Therefore,

$$\mathcal{D}_P(f, \log f) = \mu[((1-P^*)f)\log f] \geq \mathrm{Ent}_\mu f - \mathrm{Ent}_\mu P^* f \geq \delta\,\mathrm{Ent}_\mu f.$$

□

Proof of Lemma 2.3.3. The standard LSI with constant s implies entropy decay in continuous time with rate $\varrho = 2s$, since $\mathcal{D}_P(f, \log f) \geq 2\mathcal{D}_P(\sqrt{f}, \sqrt{f})$ for all $f \geq 0$. Moreover, thanks to Lemma A.1.1, we also have that entropy decay in discrete time implies MLSI. Now, suppose that the entropy decay in discrete time holds

with a constant $\delta > 0$. Pinsker's inequality says that

$$\|\delta_\sigma P^n - \pi\|_{TV}^2 \leq \frac{1}{2} H(\delta_\sigma P^n \,|\, \pi),$$

where $\delta_\sigma(\tau) = \mathbf{1}(\tau = \sigma)$ is the Dirac mass at σ. Iterating (2.3.4),

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

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

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

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

Proof of Lemma 2.3.5. The first assertion is proved in [104, Proposition 6]. The second assertion follows from the first and the simple observation that if $P = P^*$ then the LSI for (P, π) implies the LSI for $(P^* P, \pi)$ with the same constant since $P^* P = P^2 \leq P$ as quadratic forms in $L^2(\pi)$. □

Proof of Lemma 2.4.1. The identity (2.4.1) follows from (2.4.2) in the case $k = 2$ with $\Lambda_0 = \emptyset$, $\Lambda_1 = \Lambda$, $\Lambda_2 = V$. To prove (2.4.2), set $g_i = \mu_{\Lambda_i} f$, and note that $g_i = \mu_{\Lambda_i} g_{i-1}$ by (2.1.1). Therefore,

$$\begin{aligned}
\mu\left[\mathrm{Ent}_{\Lambda_k} g_0\right] &= \mu\left[g_0 \log\left(g_0/g_k\right)\right] \\
&= \sum_{i=1}^{k} \mu\left[g_{i-1} \log\left(g_{i-1}/g_i\right)\right] \\
&= \sum_{i=1}^{k} \mu\left[\mu_{\Lambda_i}\left(g_{i-1} \log\left(g_{i-1}/g_i\right)\right)\right] = \sum_{i=1}^{k} \mu\left[\mathrm{Ent}_{\Lambda_i} g_{i-1}\right].
\end{aligned}$$

□

Now we provide the proof of the Shearer inequality for product measures. This proof can be found in these notes, Lemma 4.1.

Proof of Lemma 2.4.4. We prove the Shearer inequality for the entropy functional, the proof for the variance functional is similar and thus omitted. For simplicity, we fix an ordering of the vertices of Λ and call $\Lambda = [n]$ and $\mu = \mu_\Lambda$. For any A, define $A_i = \{x \in A, x \leq i\}$ and $A_i^- = \{x \in A, x < i\}$ for any $i \in A$. Then from equation (2.4.2)

$$\mathrm{Ent}_A f = \sum_{i \in A} \mu_A \left[\mathrm{Ent}_{A_i} \mu_{A_i^-} f\right].$$

Since μ is a product measure one has $\mu_{A_i} = \mu_{\{i\}} \otimes \mu_{A_i^-}$, then

$$\mu_A \left[\mathrm{Ent}_{A_i} \mu_{A_i^-} f \right] = \mu_A \left[\mathrm{Ent}_{\{i\}} \mu_{A_i^-} f \right]. \tag{A.1.1}$$

Next we claim that

$$\mu_A \left[\mathrm{Ent}_{\{i\}} \mu_{A_i^-} f \right] \geq \mu_A \left[\mathrm{Ent}_{\{i\}} \mu_{[i-1]} f \right] \tag{A.1.2}$$

where $[i-1] = \{1, \ldots, i-1\}$ for $i > 1$ and $[i-1] = 0$ iff $i = 1$. This inequality follows from a more general statement that whenever $U, V \subset [n]$, with $U \cap V = \emptyset$ and $\mu_U \mu_V = \mu_V \mu_U$, then for all $f : \Omega \longrightarrow \mathbb{R}_+$ on has

$$\mu \left[\mathrm{Ent}_U \mu_V f \right] \leq \mu \left[\mathrm{Ent}_U f \right]. \tag{A.1.3}$$

It is immediate to see that (A.1.3) follows by applying (A.1.2) with $U = \{i\}$, $V = [i-1] \setminus A_i^-$ and f replaced by $\mu_{A_i^-} f$. To prove (A.1.3), we write

$$\begin{aligned} \mu \left[\mathrm{Ent}_U \mu_V f \right] &= \mu \left[(\mu_V f) \log \left(\frac{\mu_V f}{\mu_U \mu_V f} \right) \right] \\ &= \mu \left[(\mu_V f) \log \left(\frac{\mu_V f}{\mu_V \mu_U f} \right) \right] \\ &= \mu \left[f \log \left(\frac{\mu_V f}{\mu_V \mu_U f} \right) \right]. \end{aligned}$$

Taking $g = \log \left(\frac{\mu_V f}{\mu_V \mu_U f} \right), \nu = f \mu_U$ and observing that $\mu_U \left[e^g \right] = 1$, the variational principle (2.3.2) shows that

$$\mu_U \left[f \log \left(\frac{\mu_V f}{\mu_V \mu_U f} \right) \right] = \nu(g) \leq H(\nu | \mu_U) = \mu_U \left[f \log \left(\frac{f}{\mu_U f} \right) \right].$$

By integrating we finally get

$$\mu \left[\mathrm{Ent}_U \mu_V f \right] \leq \mu \left[\mathrm{Ent}_U f \right],$$

which implies (A.1.3). Now observe that by (2.4.2) one has

$$\mathrm{Ent}_\Lambda f = \sum_{i \in [n]} \mu \left[\mathrm{Ent}_{[i]} \mu_{[i-1]} f \right] = \sum_{i \in [n]} \mu \left[\mathrm{Ent}_{\{i\}} \mu_{[i-1]} f \right].$$

Therefore, summing over A in (A.1.1) one obtains

$$\sum_{A \subset [n]} \alpha_A \mu \left[\mathrm{Ent}_A f \right] \geq \sum_{A \subset [n]} \alpha_A \sum_{i \in A} \mu \left[\mathrm{Ent}_{\{i\}} \mu_{[i-1]} f \right] \geq \gamma(\alpha) \mathrm{Ent}_\Lambda f.$$

□

Proof of Lemma 2.4.5. From the decomposition in Lemma 2.4.1 it follows that

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

This proves (2.4.6). To prove (2.4.7) notice that by definition

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

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,

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

It remains to observe that

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

The latter estimate follows from the variational principle (2.3.2). □

Proof of Lemma 2.4.6. In view of Lemma 2.3.3 it is sufficient to prove item 2. We note that the relative entropy decay with rate δ is equivalent to the entropy contraction

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

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

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

From the decomposition in Lemma 2.4.1 it follows that

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

By Definition 2.4.3, $\sum_B \alpha_B \mu[\mathrm{Ent}_B(f)] \geq (\delta(\alpha)/C)\,\mathrm{Ent}(f)$, and therefore

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

A.2 Central limit theorem: proof of Proposition 7.1.2

In order to prove Proposition 7.1.2 we adapt to our multivariate setting some classical estimates, see e.g. [43, 112]. For the sake of clarity we give a self-contained proof. Recall the notation from Section 7.1. In particular, $|t| = \sqrt{\langle t, t\rangle}$ denotes the vector norm of $t \in \mathbb{R}^K$, and the random variable $\xi = (\xi_{i,x})$ with distribution $\mu \in \mathcal{P}(X)$ takes values in $X = \{0,1\}^K$, and has covariance matrix V_1. The next lemma only requires the nondegeneracy of μ, that is $\det(V_1) \neq 0$. The proof of Proposition 7.1.2 however requires the irreducibility of μ in order to apply Lemma 7.1.5.

Lemma A.2.1. *Let $\mu \in \mathcal{P}(X)$ be nondegenerate. For any $t \in \mathbb{R}^K$, define*

$$\varphi_N(t) := \mu\left(e^{i\langle V_N^{-1/2} t, \bar{\xi}\rangle}\right)^N,$$

where $\bar{\xi}_{i,x} := \xi_{i,x} - \mu\left[\xi_{i,x}\right]$, $V_N = NV_1$, and $L_N := \frac{1}{\sqrt{N}}\,\mu\left[\left|V_1^{-1/2}\bar{\xi}\right|^3\right]$. Then

$$\left|\varphi_N(t) - e^{-\frac{1}{2}\langle t,t\rangle}\right| \leq 16 L_N |t|^3 e^{-\frac{1}{3}\langle t,t\rangle}, \qquad |t| \leq \frac{1}{4L_N}. \tag{A.2.1}$$

Proof. We split the proof into two. First, let us suppose that $\frac{1}{4L_N} \geq |t| \geq \frac{1}{2} L_N^{-\frac{1}{3}}$. If one has

$$|\varphi_N(t)|^2 \leq e^{-\frac{2}{3}\langle t,t\rangle} \tag{A.2.2}$$

then

$$\left|\varphi_N(t) - e^{-\frac{1}{2}\langle t,t\rangle}\right| \leq |\varphi_N(t)| + e^{-\frac{1}{2}\langle t,t\rangle} \leq 2e^{-\frac{1}{3}\langle t,t\rangle} \leq 16 L_N |t|^3 e^{-\frac{1}{3}\langle t,t\rangle}.$$

We now show (A.2.2). Let $\varphi(t) := \mu\left[e^{i\langle t,\bar{\xi}\rangle}\right]$. Define $\Psi := \Psi_1 - \Psi_2$, where Ψ_1 and Ψ_2 are independent and indentically distributed as $\bar{\xi}$, so that the characteristic function of Ψ is $|\varphi(t)|^2$, and its covariance matrix is $2V_1$. Next we use

$$R(x) = e^{ix} - \left(1 + ix - \frac{x^2}{2}\right), \qquad |R(x)| \leq \min\left\{|x|^2, \frac{|x|^3}{6}\right\}, \tag{A.2.3}$$

for all $x \in \mathbb{R}$. By substituting $x = \langle t, \Psi \rangle$ in (A.2.3) and taking the expectation,

$$\begin{aligned} |\varphi(t)|^2 &= 1 - \frac{1}{2}\mu\left[(\langle t, \Psi\rangle)^2\right] + \mu\left[R\left(\langle t, \Psi\rangle\right)\right] \\ &\leq 1 - \langle t, V_1 t\rangle + \frac{1}{6}\mu\left[|\langle t, \Psi\rangle|^3\right]. \end{aligned}$$

Then, by using the fact that $(1+x)^N \leq e^{Nx}$ for all x, that $V_N^{-1/2}$ is self-adjoint, and the Cauchy-Schwarz inequality,

$$\begin{aligned} |\varphi_N(t)|^2 = \left|\varphi\left(V_N^{-1/2}t\right)\right|^{2N} &\leq \left(1 - \frac{1}{N}\langle t, t\rangle + \frac{1}{6}\mu\left[\left|\langle t, V_N^{-1/2}\Psi\rangle\right|^3\right]\right)^N \\ &\leq \left(1 - \frac{1}{N}\langle t, t\rangle + \frac{8}{6}|t|^3\mu\left[\left|V_N^{-1/2}\bar{\xi}\right|^3\right]\right)^N \\ &\leq \exp\left\{-\langle t, t\rangle + \frac{8}{6}L_N|t|^3\right\} \leq e^{-\frac{2}{3}\langle t,t\rangle}, \end{aligned}$$

thus (A.2.1) is proved if $\frac{1}{4L_N} \geq |t| \geq \frac{1}{2}L_N^{-\frac{1}{3}}$. Next, suppose that $|t| < \frac{1}{2}L_N^{-\frac{1}{3}}$. Define $\tau_N := \frac{L_N}{N}$. One has

$$\begin{aligned} \frac{1}{2} > L_N^{\frac{1}{3}}|t| > \tau_N^{\frac{1}{3}}|t| = \mu\left[\left|V_1^{-1/2}\bar{\xi}\right|^3\right]^{\frac{1}{3}}\frac{|t|}{\sqrt{N}} &\geq \mu\left[\left|V_1^{-1/2}\bar{\xi}\right|^2\right]^{\frac{1}{2}}\frac{|t|}{\sqrt{N}} \\ &\geq \frac{|t|}{\sqrt{N}}, \end{aligned} \tag{A.2.4}$$

where the last inequality follows from

$$\mu\left[\left|V_1^{-1/2}\bar{\xi}\right|^2\right]^{\frac{1}{2}} = \left(\mu\left[\langle V_1^{-1/2}\bar{\xi}, V_1^{-1/2}\bar{\xi}\rangle\right]\right)^{\frac{1}{2}} = \sqrt{K} \geq 1.$$

From (A.2.3) and the inequalities (A.2.4), using again the Cauchy-Schwarz inequality we can write

$$\begin{aligned} \left|\varphi(V_N^{-1/2}t) - 1\right| &= \left|-\frac{1}{2N}\langle t, t\rangle + \mu\left[R\left(\langle t, V_N^{-1/2}\bar{\xi}\rangle\right)\right]\right| \\ &\leq \frac{1}{2N}\langle t, t\rangle + \tau_N\frac{|t|^3}{6} < \frac{1}{8} + \frac{1}{48} < \frac{1}{4}. \end{aligned} \tag{A.2.5}$$

For all $|z| < 1$ the following inequality holds:

$$|\log(z+1) - z| \leq \frac{|z|^2}{2(1-|z|)} \tag{A.2.6}$$

Then, by using, in order, (A.2.6), (A.2.5), $|a+b|^2 \leq 2(|a|^2+|b|^2)$ and (A.2.4),

$$\left|\log\varphi(V_N^{-1/2}t) - \left(\varphi(V_N^{-1/2}t)-1\right)\right| \leq \frac{\left|\varphi(V_N^{-1/2}t)-1\right|^2}{2\left(1-\left|\varphi(V_N^{-1/2}t)-1\right|\right)} < \frac{2}{3}\left|\varphi(V_N^{-1/2}t)-1\right|^2$$

$$= \frac{2}{3}\left|-\frac{1}{2N}\langle t,t\rangle + \mu\left[R\left(\langle t, V_N^{-1/2}\bar{\xi}\rangle\right)\right]\right|^2 \leq \frac{4}{3}\left(\frac{|t|^4}{4N^2}+\frac{\tau_N^2}{36}|t|^6\right)$$
$$\leq \frac{4}{3}\left(\frac{\tau_N}{\sqrt{N}}\frac{|t|^4}{4}+\frac{\tau_N^2}{36}|t|^6\right) \leq \frac{4}{3}\left(\frac{1}{2\cdot 4}+\frac{1}{36\cdot 8}\right)\tau_N|t|^3 = \frac{37}{216}\tau_N|t|^3 < \frac{1}{5}\tau_N|t|^3.$$

By the triangular inequality we have

$$\left|\log\varphi(V_N^{-1/2}t)+\frac{1}{2N}\langle t,t\rangle\right| \leq \left|\log\varphi(V_N^{-1/2}t) - \left(\varphi(V_N^{-1/2}t)-1\right)\right| +$$
$$+\left|\left(\varphi(V_N^{-1/2}t)-1\right)+\frac{1}{2N}\langle t,t\rangle\right| \leq \frac{1}{5}\tau_N|t|^3+\frac{1}{6}\tau_N|t|^3 \leq \frac{1}{2}\tau_N|t|^3.$$

Using the inequality $|e^z-1| \leq |z|e^{|z|}$, $z\in\mathbb{C}$, and (A.2.4),

$$\left|\varphi_N(t)e^{\frac{|t|^2}{2}}-1\right| \leq \left|\log\varphi_N(t)e^{\frac{|t|^2}{2}}\right|\exp\left(\left|\log\varphi_N(t)e^{\frac{|t|^2}{2}}\right|\right)$$
$$= N\left|\log\varphi(V_N^{-1/2}t)+\frac{1}{2N}\langle t,t\rangle\right|\exp\left(N\left|\log\varphi(V_N^{-1/2}t)+\frac{1}{2N}\langle t,t\rangle\right|\right)$$

$$\leq \frac{1}{2}L_N|t|^3e^{\frac{1}{2}L_N|t|^3} \leq \frac{1}{2}L_N|t|^3e^{\frac{1}{16}} < |t|^3L_N,$$

so that (A.2.1) is proved. □

Proof of Proposition 7.1.2. Set $z_N := \frac{1}{\sqrt{N}}\,V_1^{-1/2}\left(M_N-\mu^{\otimes N}(S_N)\right)$. Using the Fourier transform as in the proof of Theorem 7.1.6 and the identity

$$\int_{\mathbb{R}^K} e^{-i\langle s,z_N\rangle-\frac{\langle s,s\rangle}{2}}ds = (2\pi)^{\frac{K}{2}}e^{-\frac{\langle z_N,z_N\rangle}{2}},$$

one has

$$\begin{aligned}
&\left|\mu^{\otimes N}\left(S_N = M_N\right) - \frac{e^{-\frac{1}{2}\langle z_N, z_N\rangle}}{(2\pi N)^{\frac{K}{2}}\sqrt{\det V_1}}\right| \\
&= \frac{1}{B_N(2\pi)^K}\left|\int_{Q_{N,K}} e^{-i\langle s, z_N\rangle}\varphi_N(s)ds - \int_{\mathbb{R}^K} e^{-i\langle s, z_N\rangle - \frac{1}{2}\langle s,s\rangle}ds\right| \qquad \text{(A.2.7)} \\
&\leq \frac{1}{B_N(2\pi)^K}\left(\int_{A_N}\left|\varphi_N(s) - e^{-\frac{1}{2}\langle s,s\rangle}\right|ds + \int_{Q_{N,K}\setminus A_N}|\varphi_N(s)|\,ds + \int_{A_N^c} e^{-\frac{1}{2}\langle s,s\rangle}ds\right),
\end{aligned}$$

where $B_N := \sqrt{\det V_N} = N^{K/2}\sqrt{\det V_1}$, $Q_{N,K} := V_N^{1/2}[-\pi,\pi]^K$, and $A_N := \{s \in \mathbb{R}^K : |s| \leq \frac{1}{4L_N}\}$. We are going to show that the three terms in the parenthesis above are bounded by $C/\sqrt{N}$.

Let us define

$$\tau := \mu\left[\left|V_1^{-1/2}\bar{\xi}\right|^3\right].$$

For the first term in (A.2.7) we use Lemma A.2.1:

$$\begin{aligned}
\int_{A_N}\left|\varphi_N(s) - e^{-\frac{<s,s>}{2}}\right|ds &\leq 16L_N\int_{A_N}|s|^3 e^{-\frac{<s,s>}{3}} \\
&\leq 16L_N\int_{\mathbb{R}^K}|s|^3 e^{-\frac{<s,s>}{3}} = \frac{C_1}{\sqrt{N}},
\end{aligned}$$

where $C_1 := 16\,\tau\int_{\mathbb{R}^K}|s|^3 e^{-\frac{<s,s>}{3}}$. For the second term in (A.2.7) we use Lemma 7.1.5:

$$\int_{Q_{N,K}\setminus A_N}|\varphi_N(s)|\,ds = B_N\int_{[-\pi,\pi)^K\setminus\mathcal{B}_\tau}|\varphi(t)|^N dt \leq B_N(2\pi)^K e^{-\frac{CN}{16\tau^2}}, \quad \text{(A.2.8)}$$

where C is defined as in that lemma and $\mathcal{B}_\tau := \left\{t \in \mathbb{R}^K : |t| \leq \frac{1}{4\tau}\right\}$. Therefore, (A.2.8) is bounded by $C_2/\sqrt{N}$ for a suitable constant C_2. Finally, to bound the last term in (A.2.7), using $L_N^{-1} \geq c\sqrt{N}$ for some constant $c > 0$, a simple estimate on the gaussian integral shows that

$$\int_{|s|>\frac{1}{4L_N}} e^{-\frac{<s,s>}{2}}ds \leq \frac{C_3}{\sqrt{N}},$$

for some constant C_3. □

Bibliography

[1] Kenneth S. Alexander. On weak mixing in lattice models. *Probability theory and related fields*, 110(4):441–471, 1998.

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

[3] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence I: modified log-Sobolev inequalities for fractionally log-concave distributions and high-temperature Ising models. *arXiv e-prints*, pages arXiv–2106, 2021.

[4] Nima Anari, Kuikui Liu, and Shayan 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.

[5] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[6] Ellen Baake, Michael Baake, and Majid Salamat. The general recombination equation in continuous time and its solution. *Discrete and Continuous Dynamical Systems*, 36(1):63–95, 2016.

[7] Franck Barthe, Dario Cordero-Erausquin, Michel Ledoux, and Bernard Maurey. Correlation and Brascamp-Lieb inequalities for Markov semigroups. *International Mathematics Research Notices*, 10:2177–2216, 2011.

[8] Giada Basile, Dario Benedetto, Emanuele Caglioti, and Lorenzo Bertini. Large deviations for a binary collision model: energy evaporation. *Mathematics in Engineering*, 5(1):1–12, 2022.

[9] Vincent Beffara and Hugo Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probability Theory and Related Fields*, 153(3-4):511–542, 2012.

[10] Lorenzo Bertini, Nicoletta Cancrini, and Filippo Cesi. The spectral gap for a Glauber-type dynamics in a continuous gas. In *Annales de l'IHP Probabilités et statistiques*, volume 38, pages 91–108, 2002.

[11] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: coupling, spectral independence, and entropy factorization. *Electronic Journal of Probability*, 27(none):1 – 42, 2022.

[12] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen–Wang dynamics on $\mathbb{Z}^d$. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1551–1564, 2021.

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

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

[15] Antonio Blanca, Reza Gheissari, and Eric Vigoda. Random-cluster dynamics in $\mathbb{Z}^2$: Rapid mixing with general boundary conditions. *The Annals of Applied Probability*, 30(1):418–459, 2020.

[16] Antonio Blanca and Alistair Sinclair. Random-cluster dynamics in $\mathbb{Z}^2$. *Probability Theory and Related Fields*, 168(3-4):821–847, 2017.

[17] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.

[18] Federico Bonetto, Alissa Geisinger, Michael Loss, and Tobias Ried. Entropy decay for the Kac evolution. *Communications in Mathematical Physics*, 363(3):847–875, 2018.

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

[20] Alexandre Bristiel and Pietro Caputo. Entropy inequalities for random walks and permutations. *Annales de l'Institut Henri Poincaré*, to appear.

[21] Russ Bubley and Martin 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.

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

[23] Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009.

[24] Pietro Caputo and Fabio Martinelli. Asymmetric diffusion and the energy gap above the 111 ground state of the quantum xxz model. *Communications in mathematical physics*, 226(2):323–375, 2002.

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

[26] Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 24, pages 691–716, 2015.

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

[28] Pietro Caputo and Daniel Parisi. Nonlinear recombinations and generalized random transpositions. *arXiv preprint arXiv:2207.04775*, 2022.

[29] Pietro Caputo and Alistair Sinclair. Entropy production in nonlinear recombination models. *Bernoulli*, 24(4B):3246–3282, 2018.

[30] Eric A. Carlen, Maria C. Carvalho, and Ester Gabetta. Central limit theorem for maxwellian molecules and truncation of the Wild expansion. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 53(3):370–397, 2000.

[31] Eric A. Carlen, Maria C. Carvalho, Jonathan Le Roux, Michael Loss, and Cédric Villani. Entropy and chaos in the Kac model. *Kinetic and Related Models*, 3(1):85–122, 2010.

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

[33] Eric A. Carlen, Elliott H. Lieb, and Michael Loss. A sharp analog of Young's inequality on S^N and related entropy inequalities. *J. Geom. Anal.*, 14(3):487–520, 2004.

[34] Patrick Cattiaux, Arnaud Guillin, and Florent Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140(1):19–40, 2008.

[35] Raphaël Cerf and Sana Louhichi. Dynamical coupling between ising and fk percolation. *Latin American Journal of Probability and Mathematical Statistics (ALEA)*, 17:23–49, 2020.

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

[37] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[38] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke 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.

[39] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. *arXiv*, 2022.

[40] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric 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.

[41] Zongchen Chen, Kuikui Liu, and Eric 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.

[42] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021.

[43] Kai Lai Chung. *A course in probability theory*. Academic press, 2001.

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

[45] Paolo Dai Pra, Anna Maria Paganoni, and Gustavo Posta. Entropy inequalities for unbounded spin systems. *The Annals of Probability*, 30(4):1959–1976, 2002.

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

[47] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[48] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):159–179, 1981.

[49] Roland L. Dobrushin and Senya B. Shlosman. Completely analytical interactions: constructive description. *Journal of Statistical Physics*, 46(5-6):983–1014, 1987.

[50] Emanuele Dolera, Ester Gabetta, and Eugenio Regazzini. Reaching the best possible rate of convergence to equilibrium for solutions of Kac's equation via central limit theorem. *The Annals of Applied Probability*, 19(1):186–209, 2009.

[51] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. Sharp phase transition for the random-cluster and potts models via decision trees. *Annals of Mathematics*, 189(1):75–99, 2019.

[52] Alain Durmus, Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. An elementary approach to uniform in time propagation of chaos. *Proceedings of the American Mathematical Society*, 148(12):5387–5398, 2020.

[53] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Dobrushin conditions and systematic scan. *Combinatorics, Probability and Computing*, 17(6):761–779, 2008.

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

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

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

[57] Amit Einav. On Villani's conjecture concerning entropy production for the Kac master equation. *Kinetic and Related Models*, 4(2):479–497, 2011.

[58] Ronen Eldan, James R. Lee, and Joseph Lehec. Transport-entropy inequalities and curvature in discrete-space Markov chains. In *A Journey Through Discrete Mathematics*, pages 391–406. 2017.

[59] Matthias Erbar, Max Fathi, and André Schlichting. Entropic curvature and convergence to equilibrium for mean-field dynamics on discrete spaces. *Latin American Journal of Probability and Mathematical Statistics*, 17(1):445, 2020.

[60] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. Ricci curvature bounds for weakly interacting markov chains. *Electronic Journal of Probability*, 22:1–23, 2017.

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

[62] Cornelius Marius Fortuin and Piet W. Kasteleyn. On the random-cluster model: I. introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[63] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field potts model. *Random Structures & Algorithms*, 54(1):82–147, 2019.

[64] Fuqing Gao and Jeremy Quastel. Exponential decay of entropy in the random transposition and bernoulli-laplace models. *The Annals of Applied Probability*, 13(4):1591–1600, 2003.

[65] Hilda Geiringer. On the probability theory of linkage in mendelian heredity. *The Annals of Mathematical Statistics*, 15(1):25–57, 1944.

[66] Sharad Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Processes and their Applications*, 114(1):51–79, 2004.

[67] Friedrich Götze, Holger Sambale, and Arthur Sinulis. Higher order concentration for functions of weakly dependent random variables. *Electronic Journal of Probability*, 24, 2019.

[68] Geoffrey R. Grimmett. *The Random-Cluster Model*, volume 333. Springer-Verlag, Berlin, 2006.

[69] F Alberto Grünbaum. Propagation of chaos for the Boltzmann equation. *Archive for Rational Mechanics and Analysis*, 42(5):323–345, 1971.

[70] Alice Guionnet and Boguslaw Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

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

[72] Heng Guo, Kaan Kara, and Ce Zhang. Layerwise systematic scan: Deep boltzmann machines and beyond. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 178–187. PMLR, 09–11 Apr 2018.

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

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

[75] Maxime Hauray and Stéphane Mischler. On kac's chaos and related problems. *Journal of Functional Analysis*, 266(10):6055–6157, 2014.

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

[77] Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *The Annals of Applied Probability*, 17(3):931–952, 2007.

[78] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.*, 46(5-6):1159–1194, 1987.

[79] Mark 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.

[80] Henry McKean Jr. Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Archive for rational mechanics and analysis*, 21(5):343–367, 1966.

[81] Henry McKean Jr. An exponential formula for solving Boltzmann's equation for a Maxwellian gas. *Journal of Combinatorial Theory*, 2(3):358–382, 1967.

[82] Mark Kac. Foundations of kinetic theory. In *Proceedings of The third Berkeley symposium on mathematical statistics and probability*, volume 3, pages 171–197, 1956.

[83] Daniel Lacker. Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *arXiv preprint arXiv:2105.02983*, 2021.

[84] Rafal Latala and Krzysztof Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, pages 147–168. Springer, 2000.

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

[86] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.

[87] David A. Levin and Yuval Peres. Markov chains and mixing times (second edition). *The Mathematical Intelligencer*, 41(1):90–91, November 2019.

[88] K. Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. In *APPROX-RANDOM*, 2021.

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

[90] Sheng Lin Lu and Horng-Tzer Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.*, 156(2):399–433, 1993.

[91] Eyal Lubetzky, Fabio Martinelli, Allan Sly, and Fabio L Toninelli. Quasi-polynomial mixing of the 2d stochastic Ising model with "plus" boundary up to criticality. *Journal of the European Mathematical Society*, 15(2):339–386, 2013.

[92] Yuri I. Lyubich, Ethan Akin, Anatolij Evgen'evič Karpov, and D. Vulis. *Mathematical structures in population genetics*, volume 22. Springer, 1992.

[93] Fabio Martinelli. *An elementary approach to finite size conditions for the exponential decay of covariances in lattice spin models*, pages 169–181. in: On Dobrushin's Way: From Probability Theory to Statistical Physics. American Mathematical Society 2000.

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

[95] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

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

[97] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Comm. Math. Phys.*, 161(3):487–514, 1994.

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

[99] Fabio Martinelli and Fabio Lucio Toninelli. On the mixing time of the 2d stochastic ising model with "plus" boundary conditions at low temperature. *Communications in Mathematical Physics*, 296(1):175–213, 2010.

[100] Servet Martínez. A probabilistic analysis of a discrete-time evolution in recombination. *Advances in Applied Mathematics*, 91:115–136, 2017.

[101] Katalin Marton. Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance. *arXiv preprint arXiv:1507.02803*, 2015.

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

[103] Carl D. Meyer. *Matrix analysis and applied linear algebra*, volume 71. SIAM, 2000.

[104] Laurent Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de markov finies. In *Séminaire de Probabilités XXXI*, pages 136–167. Springer, 1997.

[105] Laurent Miclo and Pierre Del Moral. Genealogies and increasing propagation of chaos for feynman-kac and genetic models. *The Annals of Applied Probability*, 11(4):1166–1198, 2001.

[106] Stéphane Mischler and Clément Mouhot. Kac's program in kinetic theory. *Inventiones mathematicae*, 193(1):1–147, 2013.

[107] Stéphane Mischler, Clément Mouhot, and Bernt Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probability Theory and Related Fields*, 161(1):1–59, 2015.

[108] Elchanan Mossel and Allan Sly. Exact thresholds for ising–gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.

[109] Thomas Nagylaki. *Introduction to theoretical population genetics*, volume 21. Springer Science & Business Media, 2013.

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

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

[112] Valentin V. Petrov. Sums of independent random variables. In *Sums of Independent Random Variables*. De Gruyter, 2022.

[113] Tadeusz Platkowski and Reinhard Illner. Discrete velocity models of the Boltzmann equation: a survey on the mathematical aspects of the theory. *SIAM review*, 30(2):213–255, 1988.

[114] James Gary Propp and David Bruce Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1-2):223–252, 1996.

[115] Yuval Rabini, Yuri Rabinovich, and Alistair Sinclair. A computational view of population genetics. *Random Structures & Algorithms*, 12(4):313–334, 1998.

[116] Yuri Rabinovich, Alistair Sinclair, and Avi Wigderson. Quadratic dynamical systems. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, SFCS '92, page 304–313, USA, 1992. IEEE Computer Society.

[117] Gesine Reinert and Nathan 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.

[118] Fraydoun Rezakhanlou. Propagation of chaos for particle systems associated with discrete Boltzmann equation. *Stochastic processes and their applications*, 64(1):55–72, 1996.

[119] Justin Salez. Cutoff for non-negatively curved markov chains. *Preprint, arXiv:2102.05597*, 2021.

[120] Jamil Salhi, James MacLaurin, and Salwa Toumi. On uniform propagation of chaos. *Stochastics*, 90(1):49–60, 2018.

[121] Holger Sambale and Arthur Sinulis. Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli*, 26(3):1863–1890, 2020.

[122] Daniel W. Stroock and Boguslaw Zegarliński. The logarithmic Sobolev inequality for continuous spin systems on a lattice. *J. Funct. Anal.*, 104(2):299–326, 1992.

[123] Daniel W. Stroock and Boguslaw Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992.

[124] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical review letters*, 58:86–88, 1987.

[125] Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[126] Hiroshi Tanaka. Probabilistic treatment of the boltzmann equation of maxwellian molecules. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 46(1):67–105, 1978.

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

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

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

[130] Cédric Villani. Cercignani's conjecture is sometimes true and always almost true. *Communications in mathematical physics*, 234(3):455–490, 2003.

[131] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Society, 2021.

[132] Yitong Yin Weiming Feng, Heng Guo and Chihao 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.

[133] E. Wild. On Boltzmann's equation in the kinetic theory of gases. In *Mathematical Proceedings of the cambridge Philosophical society*, volume 47, pages 602–609. Cambridge University Press, 1951.

[134] Boguslaw Zegarlinski. Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 105(1):77–111, 1992.