RAPID MIXING OF GLAUBER DYNAMICS VIA SPECTRAL INDEPENDENCE FOR ALL DEGREES

XIAOYU CHEN, WEIMING FENG, YITONG YIN, AND XINYUAN ZHANG

ABSTRACT. We prove an optimal $\Omega\left(n^{-1}\right)$ lower bound on spectral gap of the Glauber dynamics for anti-ferromagnetic two-spin systems with n vertices in the tree uniqueness regime. This spectral gap holds for all, including unbounded, maximum degree Δ. Consequently, we have the following mixing time bounds for the models satisfying the uniqueness condition with a slack $\delta \in (0, 1)$:

- $C(\delta)n^2 \log n$ mixing time for the hardcore model with fugacity $\lambda \le (1 - \delta)\lambda_c(\Delta) = (1 - \delta)\frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$;
- $C(\delta)n^2$ mixing time for the Ising model with edge activity $\beta \in \left[\frac{\Delta-2+\delta}{\Delta-\delta}, \frac{\Delta-\delta}{\Delta-2+\delta}\right]$;

where the maximum degree Δ may depend on the number of vertices n, and $C(\delta)$ depends only on δ.

Our proof is built upon the recently developed connections between the Glauber dynamics for spin systems and the high-dimensional expander walks. In particular, we prove a stronger notion of spectral independence, called the *complete spectral independence*, and use a novel Markov chain called the *field dynamics* to connect this stronger spectral independence to the rapid mixing of Glauber dynamics for all degrees.

CONTENTS

1. Introduction

Spin systems are basic graphical models for high-dimensional joint distributions expressed by pairwise interactions, and have been extensively studied in theoretical computer science, probability theory, and statistical physics. A *two-spin system* is specified on an undirected graph $G = (V, E)$ by three real parameters $\beta, \gamma, \lambda \geq 0$. Without loss of generality, one can assume that $0 \leq \beta \leq \gamma$, $\gamma > 0$ and $\lambda > 0$. Each *configuration* $\sigma \in \{-1, +1\}^V$ assigns every vertex $v \in V$ one of the two *spin states* from $\{-1, +1\}$. A probability distribution μ over all configurations $\sigma \in \{-1, +1\}^V$, called *Gibbs distribution*, is defined as:

$$\mu(\sigma) = \frac{1}{Z} \beta^{m_{+1}(\sigma)} \gamma^{m_{-1}(\sigma)} \lambda^{n_{+1}(\sigma)},$$

where $m_i(\sigma) \triangleq |\{(u, v) \in E \mid \sigma_u = \sigma_v = i\}|$ denotes the number of i-monochromatic edges for $i \in \{-1, +1\}$ and $n_{+1}(\sigma) \triangleq |\{v \in V \mid \sigma_v = +1\}|$ denotes the number of $+1$-spin vertices, and the normalizing factor

$$Z = \sum_{\sigma \in \{-1, +1\}^V} \beta^{m_{+1}(\sigma)} \gamma^{m_{-1}(\sigma)} \lambda^{n_{+1}(\sigma)}$$

gives the *partition function*. The two-spin system is called *ferromagnetic* if $\beta\gamma > 1$ and is called *anti-ferromagnetic* if $\beta\gamma < 1$. In particular, two extensively studied classes of two-spin systems are:

- The *hardcore model* with *fugacity* λ, which corresponds to two-spin systems with $\beta = 0$ and $\gamma = 1$. Every configuration σ with $\mu(\sigma) > 0$ corresponds to an independent set I_σ in G, and $\mu(\sigma) \propto \lambda^{|I_\sigma|}$.
- The *Ising model* with *edge activity* β and *external field* λ, which corresponds to two-spin systems with $\beta = \gamma$. The Gibbs distribution becomes $\mu(\sigma) \propto \beta^{m(\sigma)} \lambda^{n_{+1}(\sigma)}$, where $m(\sigma) = m_{-1}(\sigma) + m_{+1}(\sigma)$ gives the number of monochromatic edges.

A phenomenon that has drawn considerable attention of two-spin systems is the computational phase transition for sampling. In a seminal work [?], by exploiting a phase transition property based on decay of correlation, known as the *spatial mixing*, Weitz shew that sampling from the Gibbs distribution μ of the hardcore model with fugacity $\lambda < \lambda_c(\Delta)$ on any n-vertex graph of bounded maximum degree Δ is tractable in time $n^{O(\log \Delta)}$. Here the critical threshold $\lambda_c(\Delta) \triangleq \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ is a famous threshold for the uniqueness/non-uniqueness phase transition for the hardcore model on infinite Δ-regular tree [?], beyond which, i.e. when $\lambda > \lambda_c(\Delta)$, the infinite-volume Gibbs measure on Δ-regular tree is not uniquely defined, and approximately sampling in the hardcore model on graphs with bounded maximum degree Δ becomes computationally intractable [? ? ?].

This sharp computational phase transition was extended to all anti-ferromagnetic two-spin systems of maximum degree Δ specified by (β, γ, λ) that is *up-to-Δ unique* [? ? ?], which corresponds to the uniqueness condition for infinite regular trees up to degree Δ. And if (β, γ, λ) lies outside this regime, the sampling problem becomes computationally intractable [? ?]. Similar bounds were also achieved by another family of critical algorithms based on polynomial interpolation approach for approximating non-vanishing polynomials [? ? ?]. A glaring issue with both these families of critical algorithms is the expensive time cost, usually in a form of $n^{O(\log \Delta)}$, which is due to enumerating $O(\log n)$-sized local structures in these algorithms. Such time complexity grows super-polynomially in the size n of the model when the maximum degree Δ is unbounded.

A major open question is whether sampling in spin systems is always tractable for the class of instances within the uniqueness regime, which should by no means restrict to the graphical models with universally bounded max-degrees. In particular, we wonder whether such *fixed-parameter tractable* algorithms exist:

Question 1. Let $\delta \in (0, 1)$ be an arbitrary gap. Can we sample approximately in the hardcore model on any n-vertex graph G of maximum degree Δ_G with fugacity $\lambda \leq (1 - \delta)\lambda_c(\Delta_G)$ in time $f(\delta) \cdot \text{poly}(n)$?

For anti-ferromagnetic two-spin systems, the condition $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ is generalized by the regime of (β, γ, λ) that is *up-to-Δ unique with gap δ*, which corresponds to the interior (determined by the gap δ) of the regime for the uniqueness condition on infinite regular trees up to degree Δ.

There is a canonical Markov chain based algorithm for sampling from Gibbs distributions known as the *Glauber dynamics* (a.k.a *heat bath*, *Gibbs sampling*). The Glauber dynamics for a joint distribution μ of variables from V is a Markov chain $(X_t)_{t \geq 0}$ on space $\Omega(\mu)$, where $\Omega(\mu)$ denotes the the support of μ. At the t-th step, the rule for updating X_t is:

- pick a $v \in V$ uniformly at random;
- update $X_t(v)$ according to μ projected onto v given the boundary condition $X_t(V \setminus \{v\})$.

The chain is stationary and reversible at μ [?]. The rate of convergence is given by the *mixing time*:

$$\forall 0 < \varepsilon < 1, \quad T_{\text{mix}}(\varepsilon) \triangleq \max_{X_0 \in \Omega(\mu)} \min \{t \mid d_{\text{TV}}(X_t, \mu) \leq \varepsilon\},$$

where $d_{\text{TV}}(X_t, \mu)$ denotes the total variation distance between the distribution of X_t and μ.

Let P denote the transition matrix of the Glauber dynamics, which is positive semidefinite [?]. Then P has non-negative real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\Omega(\mu)|} \geq 0$. The *spectral gap* is defined by $\lambda_{\text{gap}}(P) \triangleq 1 - \lambda_2$. Given the spectral gap, the mixing time is bounded as:

$$(1) \qquad\qquad T_{\text{mix}}(\varepsilon) \leq \frac{1}{\lambda_{\text{gap}}(P)} \log\left(\frac{1}{\varepsilon \cdot \mu_{\text{min}}}\right), \quad \text{where } \mu_{\text{min}} \triangleq \min_{\sigma \in \Omega(\mu)} \mu(\sigma).$$

It was widely speculated that for a wide range of spin systems, the computational phase transition for sampling is captured by the rapid mixing of Glauber dynamics

Using coupling based techniques, one can obtain $O(n \log n)$ mixing time bounds. However, previous critical results using coupling methods held either by assuming girth lower bounds [? ?] or by assuming ferromagnetism and bounded maximum degree [?].

In a series of breakthrough works [? ? ?], the Glauber dynamics was interpreted as a higher order random walk on simplicial complexes, and techniques for high-dimensional expander walks were applied to analyze its mixing. In particular, Alev and Lau [?] gave a sharp "local-to-global" argument for lifting the spectral expansions from a local down-up walk to the high-dimensional expander walk. In a seminal work [?], Anari, Liu and Oveis Gharan formulated a key concept, called the *spectral independence*, which is measured by the spectral radius of the influence matrix, in which each entry $(u, v) \in V^2$ gives the influence of u's spin on the marginal probability at v in the Gibbs distribution. For product distributions, this value is 0. Intuitively, it measures how variables are independent of each other in a joint distribution. The concept intrinsically connects the spatial mixing properties with the local spectral expansions of high-dimensional walks. Then by utilizing the "local-to-global" result of [?], Anari, Liu and Oveis Gharan [?] proved an $n^{\exp(O(1/\delta))}$ mixing time bound for the hardcore model with fugacity $\lambda \leq (1-\delta)\lambda_c(\Delta)$. In [?], Chen, Liu and Vigoda proved tight bounds on spectral independence from spatial mixing properties, and consequently gave an improved $n^{O(1/\delta)}$ mixing time bound for all anti-ferromagnetic two-spin systems that are up-to-Δ unique with gap δ, that is, when the uniqueness condition is satisfied with a slack δ.

This series of breakthroughs culminated in a $C(\delta, \Delta)n \log n$ mixing time bound for all anti-ferromagnetic two-spin systems satisfying the up-to-Δ uniqueness with gap δ by Chen, Liu and Vigoda [?], which was obtained by an ingenious "local-to-global" argument for the relative entropy decay (modified log-Sobolev constant) in a multi-level down-up walk that corresponds to heat-bath block dynamics. It gave an optimal $O(n \log n)$ mixing time when the max-degree $\Delta = O(1)$, although in general the bound may grow like $\Delta^{O(\Delta^2/\delta)}n \log n$. As observed in a very recent work [?], a variance-decay variant of [?] could in fact imply a $\tilde{O}(\Delta^{O(1/\delta)}n^2)$ mixing time. All in all, the previous best bound on the mixing time for classes of near-critical instances without degree or girth restriction is $n^{O(1/\delta)}$.

Now look back to the problems like Question 1. The major unresolved instances are the ones that have very large degree (e.g. $\Delta = n^{\Omega(1)}$) and contain many small cycles. Indeed for such instances, it is unknown whether local Markov chains as Glauber dynamics should mix rapidly, or they should be hard instances that prevent sampling from being tractable while only the sub-criticality of the system is fixed.

1.1. Results for two-spin systems.

We show that large degree and small cycles do not slow down mixing. In particular, we give an optimal lower bound for the spectral gap of the Glauber dynamics for all anti-ferromagnetic two-spin systems within the uniqueness regime.

Theorem 1.1. *For all $\delta \in (0, 1)$, there exists a finite $C(\delta) > 0$ such that for every anti-ferromagnetic two-spin system on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ specified by (β, γ, λ) that is up-to-Δ unique with gap δ, the spectral gap of the Glauber dynamics is*

$$\lambda_{\text{gap}} \geq \frac{1}{C(\delta)n}.$$

Remark 1.2. The constant $C(\delta)$ in the theorem depends only on the gap δ, and is independent of the maximum degree Δ and the parameters (β, γ, λ). This gives an optimal spectral gap $\Omega(n^{-1})$ for all anti-ferromagnetic two-spin systems satisfying the uniqueness condition with a constant gap, regardless of the maximum degree Δ. More precisely, $C(\delta)$ is bounded as $C(\delta) = (\frac{1}{\delta})^{O(1/\delta)}$, and for the hardcore or Ising models, $C(\delta)$ can be further improved to $C(\delta) = \exp(O(\frac{1}{\delta}))$.

Due to the well known relation between the spectral gap and the mixing time in (1), Theorem 1.1 implies the mixing time bound for the same class of spin systems:

$$(2) \qquad T_{\text{mix}}(\varepsilon) \leq C(\delta)n \left(n \log \left(\lambda + \frac{1}{\lambda} \right) + n\Delta \log \alpha + \log \frac{1}{\varepsilon} \right),$$

where $C(\delta) = (\frac{1}{\delta})^{O(1/\delta)}$ is the same factor as in Theorem 1.1, $\alpha = (\frac{1}{\beta} + 2)$ when $\beta > 0$ and $\alpha = (\gamma + \frac{1}{\gamma} + 2)$ when $\beta = 0$. Therefore, the mixing time of the Glauber dynamics for anti-ferromagnetic 2-spin systems within the uniqueness regime is always bounded by $n^{3+o(1)}$, even when the maximum degree Δ or the parameters (β, γ, λ) may depend on n, as long as the positive $\beta, \gamma, \lambda \in \exp(n^{o(1)}) \cap \exp(-n^{o(1)})$.

For the hardcore model and the Ising model, we have even better bounds.

Theorem 1.3. *For all $\delta \in (0, 1)$, there exists a $C(\delta) = \exp(O(\frac{1}{\delta}))$ such that for every hardcore model on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ and with fugacity $\lambda \leq (1 - \delta)\lambda_c(\Delta)$, the spectral gap and the mixing time of the Glauber dynamics are respectively bounded as*

$$\lambda_{\text{gap}} \geq \frac{1}{C(\delta)n} \quad \text{and} \quad T_{\text{mix}}(\varepsilon) \leq C(\delta)n \left(n \log \Delta + \log \frac{1}{\varepsilon} \right).$$

As far as we know, this is the first fixed-parameter-tractable (FPT) result for the hardcore model within the uniqueness regime, where the gap δ to the critical threshold is the only fixed parameter. Previously, for the same hardcore uniqueness regime, [?] gave an $n^{O(1/\delta)}$ and [?] gave a $\Delta^{O(\Delta^2/\delta)} n \log n$ upper bounds on the mixing time. Here we give a $\tilde{O}(n^2)$ mixing time bound, which is always bounded by a polynomial of absolute constant degree, no matter how large the graph maximum degree Δ is. And unlike the mixing results in [? ? ?], our result needs not to assume girth lower bound.

Theorem 1.4. *For all $\delta \in (0, 1)$, there exists a $C(\delta) = \exp(O(\frac{1}{\delta}))$ such that for every Ising model on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ and with edge activity $\beta \in \left[\frac{\Delta-2+\delta}{\Delta-\delta}, \frac{\Delta-\delta}{\Delta-2+\delta} \right]$ and external field $\lambda > 0$, the spectral gap and the mixing time of the Glauber dynamics are respectively bounded as*

$$\lambda_{\text{gap}} \geq \frac{1}{C(\delta)n} \quad \text{and} \quad T_{\text{mix}}(\varepsilon) \leq C(\delta)n \left(n + \log \frac{1}{\varepsilon} \right).$$

Compared to [?], which gave $\Delta^{O(1/\delta)} n \log n$ mixing time bound for the same Ising uniqueness regime, and [?], which gave $\exp(\Delta^{O(1/\delta)}) n \log n$ mixing time bound for the ferromagnetic half of this regime, our mixing time is bounded by $O(n^2)$ for all, including the unbounded, maximum degrees Δ, while previously the best known mixing time upper bound when Δ is unbounded is $n^{O(1/\delta)}$ [? ?].

1.2. Results for spectrally independent joint distributions. Let V be a set of Boolean random variables, and μ a distribution over $\{-1, +1\}^V$. We use $\Omega(\mu)$ to denote the support of μ. A configuration $\tau \in \{-1, +1\}^V$ is *feasible* if $\tau \in \Omega(\mu)$. For any subset $\Lambda \subseteq V$, let μ_Λ denote the marginal distribution on Λ projected from μ. A partial configuration $\tau \in \{-1, +1\}^\Lambda$, where $\Lambda \subseteq V$, is feasible if $\tau \in \Omega(\mu_\Lambda)$. For $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, we use μ^{σ_Λ} to denote the distribution over $\{-1, +1\}^V$ induced by μ conditioned on the configuration on Λ being fixed as σ_Λ. Formally:

$$\forall \tau \in \{-1, +1\}^V, \quad \mu^{\sigma_\Lambda}(\tau) = \mathbf{Pr}_{X \sim \mu}\left[X = \tau \mid X_\Lambda = \sigma_\Lambda\right].$$

For $S \subseteq V$, we use $\mu_S^{\sigma_\Lambda}$ to denote the marginal distribution on S projected from μ^{σ_Λ}, and write $\mu_v^{\sigma_\Lambda} = \mu_{\{v\}}^{\sigma_\Lambda}$.

Definition 1.5 (spectral independence). Let μ be a distribution over $\{-1, +1\}^V$. For any $\Lambda \subseteq V$, any $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, the *influence matrix* $\Psi_\mu^{\sigma_\Lambda} \in \mathbb{R}_{\geq 0}^{(V \setminus \Lambda) \times (V \setminus \Lambda)}$ is defined as: $\Psi_\mu^{\sigma_\Lambda}(u, u) = 0$ for all $u \in V \setminus \Lambda$; and for all distinct $u, v \in V \setminus \Lambda$,

$$\Psi_\mu^{\sigma_\Lambda}(u, v) \triangleq \max_{c, c' \in \Omega(\mu_u^{\sigma_\Lambda})} d_{\mathrm{TV}}\left(\mu_v^{\sigma_\Lambda, u \leftarrow c}, \mu_v^{\sigma_\Lambda, u \leftarrow c'}\right),$$

where $\mu_v^{\sigma_\Lambda, u \leftarrow c}$ denotes the marginal distribution $\mu_v^{\sigma_\Lambda}$ further conditioned on the value of v being fixed as c, and $d_{\mathrm{TV}}(\cdot, \cdot)$ denotes the total variation distance.

Let $\eta > 0$. The distribution μ is said to be *η-spectrally independent* if for any $\Lambda \subseteq V$, any $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, the spectral radius of the influence matrix $\Psi_\mu^{\sigma_\Lambda}$ has

$$\rho\left(\Psi_\mu^{\sigma_\Lambda}\right) \leq \eta. \tag{3}$$

Remark 1.6. The above definition is an alternative to the original definition of the spectral independence due to Anari, Liu and Oveis Gharan [?], who defined the notion using the signed influence matrix $I_\mu^{\sigma_\Lambda}(u, v) \triangleq \mu_v^{\sigma_\Lambda, u \leftarrow +1}(+1) - \mu_v^{\sigma_\Lambda, u \leftarrow -1}(+1)$. Here for some technical reasons (explained in Section 2.2), we adopt the notion of spectral independence proposed in [?] defined using the absolute influence matrix. The two definitions are morally equivalent, and in spin systems they can both be implied by the spatial mixing property [? ? ? ?].

In a seminal work [?], Anari, Liu and Oveis Gharan introduced the notion of spectral independence and proved a $n^{O(\eta)}$ mixing time bound for η-spectrally independent joint distributions of Boolean variables. For Gibbs distributions specified by spin systems, in a recent major breakthrough [?] Chen, Liu and Vigoda proved a $C(\delta, \Delta)n \log n$ upper bound for the mixing time of Glauber dynamics after establishing a tight $O(1/\delta)$-spectral independence [?]. This mixing time bound is remarkably optimal when the max-degree Δ is bounded by a constant, however in general the $C(\delta, \Delta)$ factor can grow as fast as $\Delta^{O(\Delta^2/\delta)}$.

In this paper, we give a novel connection between the spectral independence and rapid mixing of the Glauber dynamics, which can give us optimal spectral gap bounds without restricting the maximum degree of the graphical model. To formally state our result, we introduce the following notions.

Definition 1.7 (magnetizing a joint distribution with local fields). Let μ be a distribution over $\{-1, +1\}^V$. Let $\boldsymbol{\phi} = (\phi_v)_{v \in V}$, where each $\phi_v > 0$ specifies a *local field* at v. Denote by $\mu^{(\boldsymbol{\phi})}$ the distribution obtained from imposing the local fields $\boldsymbol{\phi}$ onto μ. Formally, $\pi = \mu^{(\boldsymbol{\phi})}$ is a distribution over $\{-1, +1\}^V$ such that:

$$\forall \sigma \in \{-1, +1\}^V, \quad \pi(\sigma) \propto \mu(\sigma) \prod_{v \in V : \sigma_v = +1} \phi_v.$$

In particular, if $\boldsymbol{\phi}$ is a constant vector with $\phi_v = \theta$ for all $v \in V$ for some scalar $\theta > 0$, we write $\mu^{(\theta)} = \mu^{(\boldsymbol{\phi})}$.

After magnetizing with local fields $\boldsymbol{\phi}$, each variable $v \in V$ is locally biased towards $+1$ if $\phi_v > 1$ or towards -1 if $\phi_v < 1$. Without loss of generality, we consider only the case with $\phi_v \leq 1$, which can cover the $\phi_v > 1$ case by flipping the roles of -1 and $+1$ for variable v.

Definition 1.8 (complete spectral independence). *Let $\eta > 0$. A distribution μ over $\{-1, +1\}^V$ is said to be completely η-spectrally independent if $\mu^{(\phi)}$ is η-spectrally independent for all $\phi \in (0, 1]^V$.*

The notion captures a desirable situation: when establishing the spectral independence, one usually proves stronger results, so that the spectral independence remains to hold for all smaller local fields. For instance, in the hardcore model, decreasing λ would only make the model more spectrally independent. For general anti-ferromagnetic 2-spin systems, this becomes more complicated, nevertheless, due to Theorem 2.10, the same holds after properly flipping the roles of -1 and +1 for variables.

Let μ be a distribution over $\{-1, +1\}^V$. We use $\lambda_{\text{gap}}^{\text{GD}}(\mu)$ to denote the spectral gap of the Glauber dynamics for μ. We further consider the spectral gap up to worst-case pinning, which is defined as:

$$(4) \qquad \lambda_{\text{min-gap}}^{\text{GD}}(\mu) \triangleq \min_{\Lambda \subseteq V, \sigma_\Lambda \in \Omega(\mu_\Lambda)} \lambda_{\text{gap}}^{\text{GD}}(\mu^{\sigma_\Lambda}),$$

where $\lambda_{\text{gap}}^{\text{GD}}(\mu^{\sigma_\Lambda})$ gives the spectral gap of Glauber dynamics for the joint distribution μ^{σ_Λ} over $\{-1, +1\}^V$ conditioned on σ_Λ, with convention that $\lambda_{\text{gap}}^{\text{GD}}(\mu^{\sigma_\Lambda}) = 1$ if μ^{σ_Λ} has trivial support with $|\Omega(\mu^{\sigma_\Lambda})| = 1$.

Theorem 1.9 (main technical theorem). *Let μ be a distribution over $\{-1, +1\}^V$, and $\eta > 0$. If μ is completely η-spectrally independent, then for all $\theta \in (0, 1)$,*

$$\lambda_{\text{gap}}^{\text{GD}}(\mu) \geq \left(\frac{\theta}{2}\right)^{2\eta+7} \cdot \lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(\theta)}\right).$$

The above theorem can be thought as a boosting theorem. After magnetizing with external field θ, the original near-critical μ is transformed to a much easier $\mu^{(\theta)}$ whose spectral gap is either known or easy to bound using standard approaches. Then Theorem 1.9 effectively boosts the mixing result for the easier distribution $\mu^{(\theta)}$ up to the near-critical regime, with a $\theta^{O(\eta)}$ overhead which is determined by both the degree of spectral independence and the distance between the critical threshold and the easier regime.

Example 1.10 (hardcore model). *A generalization of the analyses in [?] shows that the Gibbs distribution μ of the hardcore model with fugacity $\lambda \leq (1-\delta)\lambda_c(\Delta)$ is completely $O(\frac{1}{\delta})$-spectrally independent (formally stated in Theorem 2.10). We then choose $\theta = \frac{1}{25}$, so that $\mu^{(\theta)}$ corresponds to the hardcore model with a smaller fugacity*

$$\theta\lambda = \frac{\lambda}{25} < \frac{(\Delta-1)^{\Delta-1}}{25(\Delta-2)^\Delta} \leq \frac{1}{2\Delta},$$

where the last inequality holds for all $\Delta \geq 3$. In that regime, there is a coupling of the Glauber dynamics that decays step-wise, and consequently (see [?] and [? , ch. 13.1]), the spectral gap $\lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(\theta)}\right) \geq \frac{1}{2n}$. Then by Theorem 1.9, the Gibbs distribution μ of the hardcore model with fugacity $\lambda \leq (1-\delta)\lambda_c(\Delta)$ has

$$\lambda_{\text{gap}}^{\text{GD}}(\mu) \geq \theta^{O(\frac{1}{\delta})} \cdot \lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(\theta)}\right) \geq \frac{1}{\exp(O(\frac{1}{\delta})) \cdot n}.$$

Assuming $\frac{1}{2\Delta} \leq \lambda < \lambda_c(\Delta)$ (the $\lambda < \frac{1}{2\Delta}$ case is dealt by coupling), it holds that $\mu_{\min} \geq (\frac{1}{8\Delta})^n$. Thus by (1),

$$T_{\text{mix}}(\varepsilon) \leq \frac{1}{\lambda_{\text{gap}}^{\text{GD}}(\mu)} \log\left(\frac{1}{\varepsilon\mu_{\min}}\right) \leq \exp(O(1/\delta)) \cdot n\left(n\log\Delta + \log\frac{1}{\varepsilon}\right).$$

Remark 1.11. For general anti-ferromagnetic two-spin systems, similar arguments apply, except that naively decreasing the external field λ might not work. Nevertheless, we observe that for every variable there always exists a good direction pointing to either -1 or $+1$ such that biasing the local field of the variable towards the good direction may only make the model easier. This is formally stated in Theorem 2.10. Therefore, we can flip the roles of -1 and +1 for those variables whose good direction is pointing to +1, which results in a new Gibbs distribution ν that is isomorphic to μ, such that the spectral independence for anti-ferromagnetic two-spin systems proved in [?] in fact means the complete spectral independence of ν. Then the same argument as the hardcore model can apply. The details are given in Section 8.

2. The Field Dynamics and Proofs Outline

2.1. **The field dynamics.** We introduce a novel Markov chain called *the field dynamics*, which plays a key role in the proof of our main theorem (Theorem 1.9) and may be of independent interests.

Given a distribution μ over $\{-1, +1\}^V$ and a scalar $\theta \in (0, 1)$, the field dynamics for μ with parameter θ, denoted by P_θ^{FD}, is a Markov chain on space $\Omega(\mu)$ defined as follows.

The Field Dynamics

The initial state is an arbitrary feasible configuration $\sigma \in \Omega(\mu)$;
The rule for updating a configuration $\sigma \in \Omega(\mu)$ is:

   (1) generate a random $S \subseteq V$ by independently selecting each $v \in V$ into S with probability

$$p_v \triangleq \begin{cases} 1 & \text{if } \sigma_v = -1, \\ \theta & \text{if } \sigma_v = +1; \end{cases}$$

   (2) replace σ_S by a random partial configuration sampled according to $\pi_S^{\sigma_{V \setminus S}}$, where $\pi = \mu^{(\theta)}$.

The field dynamics can be thought as an adaptive variant of the heat-bath block dynamics. In each step, a block S is randomly generated adaptively to the current configuration σ. And the configuration on S is resampled given the boundary condition $\sigma_{V \setminus S}$ according to a properly biased distribution $\pi = \mu^{(\theta)}$ for canceling with the bias introduced in the adaptive construction of block S.

For instance, on the hardcore model, the field dynamics gives us the following novel Markov chain:

Example 2.1 (the field dynamics for the hardcore model). *Let μ be the Gibbs distribution of the hardcore model on graph $G = (V, E)$ with fugacity λ. The rule for updating the current $\sigma \in \Omega(\mu)$:*

- *each occupied vertex v (with $\sigma_v = +1$) is picked independently with probability $1 - \theta$; all picked vertices and their neighbors are removed from V and let S° denote the set of remaining vertices;*
- *resample σ_{S° according to the Gibbs distribution π of the hardcore model on $G[S^\circ]$ with fugacity $\theta\lambda$.*

The next proposition shows that the field dynamics P_θ^{FD} for μ has the stationary distribution μ, and is always ergodic and reversible (hence having real eigenvalues). The proposition is proved in Section 5.1.

Proposition 2.2. *For every distribution μ over $\{-1, +1\}^V$, and for all $\theta \in (0, 1)$, the field dynamics P_θ^{FD} for μ is irreducible, aperiodic and reversible with respect to μ.*

Let $\lambda_{\mathrm{gap}}^{\mathsf{FD}}(\mu, \theta)$ denote the spectral gap of the field dynamics P_θ^{FD} for μ. We state two key lemmas for the spectral gap of the field dynamics. The first is a *mixing lemma* that guarantees the fast mixing of the field dynamics for μ assuming the complete spectral independence of μ.

Lemma 2.3 (field dynamics - mixing lemma). *Let μ be a distribution over $\{-1, +1\}^V$ and $\eta > 0$. If μ is completely η-spectrally independent, then for all $\theta \in (0, 1)$,*

$$\lambda_{\mathrm{gap}}^{\mathsf{FD}}(\mu, \theta) \geq \left(\frac{\theta}{2}\right)^{2\eta+7}.$$

On the other hand, the following *comparison lemma* relates the spectral gap of the field dynamics to that of the Glauber dynamics, which holds universally for all joint distributions μ with Boolean domain.

Lemma 2.4 (field dynamics - comparison lemma). *Let μ be a distribution over $\{-1, +1\}^V$. For all $\theta \in (0, 1)$,*

$$\lambda_{\mathrm{gap}}^{\mathsf{GD}}(\mu) \geq \lambda_{\mathrm{gap}}^{\mathsf{FD}}(\mu, \theta) \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathsf{GD}}\left(\mu^{(\theta)}\right).$$

Our main technical result (Theorem 1.9) follows immediately from Lemma 2.3 and Lemma 2.4. In the next, we give outlines of the proofs of these two key lemmas.

2.2. **Mixing of field dynamics via spectral independence**. We now give an outline of our proof of the mixing lemma for the field dynamics (Lemma 2.3). A key observation is that the field dynamics is in fact a limiting instance for the uniform block dynamics.

Let μ be a distribution over $\{-1, +1\}^V$. The *uniform ℓ-block dynamics* for μ is basically the heat-bath block dynamics for μ where in each step a subset of vertices of size ℓ is chosen uniformly at random and gets updated. Specifically, the Markov chain is on space $\Omega(\mu)$ and in each step a subset $S \subseteq V$ of ℓ variable is chosen uniformly at random and the configuration on S is updated according to μ conditional on the current configuration on $V \setminus S$. In [?], Chen Liu and Vigoda proved the following theorem for the rapid mixing of the uniform block dynamics from spectral independence.

Theorem 2.5 ([?]). *Let μ be a distribution over $\{-1, +1\}^V$, $n = |V|$ and $\eta > 0$. If μ is η-spectrally independent, then for any $2\lceil \eta \rceil \leq \ell \leq n$, the uniform ℓ-block dynamics P_ℓ for μ has $\lambda_{\mathrm{gap}}(P_\ell) \geq \left(\frac{\ell}{2n}\right)^{2\lceil \eta \rceil + 1}$.*

This is a general result that holds for all joint distributions, and imposes no restriction on the maximum degree of the underlying model. In fact, the entropy (modified log-Sobolev) variant of Theorem 2.5 proved in [?] required small degrees. The statement in Theorem 2.5 is a consequence to the "local-to-global" argument for the variance contraction [? , Theorem A.9]. For completeness, we provide a proof of Theorem 2.5 in Appendix A using our notion of spectral independence with absolute influence matrix.

Previously, rapid mixing was established for the Glauber dynamics by directly comparing it with the uniform block dynamics [?], which resulted in a super-polynomial reliance on the max-degree.

Here, we describe a novel way to utilize the rapid mixing of uniform block dynamics.

Definition 2.6 (k-transformation). Let μ be a distribution over $\{-1, +1\}^V$ and $k \geq 1$ an integer. The k-*transformation* of μ, denoted by $\mu_k = \mathrm{Trans}(\mu, k)$, is a distribution over $\{-1, +1\}^{V \times [k]}$ defined as follows.

Let $X \sim \mu$. Then $\mu_k = \mathrm{Trans}(\mu, k)$ is the distribution of $Y \in \{-1, +1\}^{V \times [k]}$ constructed as follows:

- if $X_v = -1$, then $Y_{(v,i)} = -1$ for all $i \in [k]$;
- if $X_v = +1$, then $Y_{(v,i^*)} = +1$ and $Y_{(v,i)} = -1$ for all $i \in [k] \setminus \{i^*\}$, where i^* is chosen from $[k]$ uniformly and independently at random.

We use $P_{k,\ell}$ to denote the uniform ℓ-block dynamics for μ_k, and $\lambda_{\mathrm{gap}}(P_{k,\ell})$ its spectral gap. Recall that $\lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu, \theta)$ denotes the spectral gap of the field dynamics P_θ^{FD} for μ with parameter $\theta \in (0, 1)$.

One of our key discoveries is that the field dynamics P_θ^{FD} for μ is a limiting instance for the uniform $\lceil \theta k n \rceil$-block dynamics running on the k-transformations of μ.

Lemma 2.7. *For all $\theta \in (0, 1)$, it holds that*

$$\lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu, \theta) \geq \limsup_{k \to \infty} \lambda_{\mathrm{gap}}\left(P_{k, \lceil \theta k n \rceil}\right).$$

The lemma is proved in Sections 5.2 and 6. In a high level, it is proved as follows: Each feasible $\sigma \in \Omega(\mu_k)$ in the k-transformed distribution μ_k can be naturally projected back to a feasible $\sigma^\star \in \Omega(\mu)$ in the original μ, where $\sigma_v^\star$ indicates whether $\sigma_{(v,i)} = +1$ for some $i \in [k]$. This also naturally projects the block dynamics $P_{k, \lceil \theta k n \rceil}$ to a chain on $\Omega(\mu)$ that essentially preserves the spectral gap. And more crucially, the projected chain gives an entry-wise approximation of the field dynamics P_θ^{FD}.

Next, we prove a Chen-Liu-Vigoda theorem (Theorem 2.5) for the k-transformed distributions.

Lemma 2.8. *Let μ be a distribution over $\{-1, +1\}^V$, $n = |V|$ and $\eta > 0$. If μ is completely η-spectrally independent, then for any integers $k \geq 1$ and $2(\eta + 3) \leq \ell \leq kn$, it holds that $\lambda_{\mathrm{gap}}(P_{k,\ell}) \geq \left(\frac{\ell}{2kn}\right)^{2\eta + 7}$.*

The lemma is proved in Section 7. Note that Lemma 2.8 assumes only the spectral independence for μ, while the block dynamics is for μ_k. We show that assuming the *complete* spectral independence of μ, the k-transformation may only cause a constant additive overhead to the spectral independence. This is the

part where we are technically more convenient to work with absolute influence matrices rather than the signed ones, because of the monotonicity of spectral radius that holds for nonnegative matrices.

Lemma 2.7 and Lemma 2.8 together suffice to prove the mixing lemma for the field dynamics (Lemma 2.3).

Proof of Lemma 2.3. Let $\theta \in (0, 1)$. Let $(A_k)_{k \in \mathbb{N}}$ and $(B_k)_{k \in \mathbb{N}}$ be constructed as:

$$A_k = \lambda_{\text{gap}} \left(P_{k, \lceil \theta k n \rceil} \right) \quad \text{and} \quad B_k = \begin{cases} \left(\frac{\lceil \theta k n \rceil}{2kn} \right)^{2\eta+7} & \text{if } k \geq \frac{2(\eta+3)}{\theta n}, \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 2.8, $2 \geq A_k \geq B_k \geq 0$ for all $k \geq 1$. We have that $\limsup_{k \to \infty} A_k \geq \limsup_{k \to \infty} B_k$.

Since $\lim_{k \to \infty} B_k = \left(\frac{\theta}{2} \right)^{2\eta+7}$, by Lemma 2.7, we have

$$\lambda_{\text{gap}}^{\text{FD}}(\mu, \theta) \geq \limsup_{k \to \infty} \lambda_{\text{gap}} \left(P_{k, \lceil \theta k n \rceil} \right) \geq \left(\frac{\theta}{2} \right)^{2\eta+7}. \qquad \square$$

2.3. **Comparing Glauber dynamics with field dynamics**. With the rapid mixing result we just proved for the field dynamics, we already have an efficient sampling algorithm for distribution μ. Suppose that we try to implement the field dynamics. The nontrivial operation in each step of the field dynamics, namely the step of resampling from the marginal distribution induced by $\pi = \mu^{(\theta)}$ on a randomly generated block, can be simulated by running a Glauber dynamics for π on the block with boundary condition, which is known to be rapidly mixing because $\pi = \mu^{(\theta)}$ lies in a much easier regime!

Example 2.9 (a hardcore sampler). *For the hardcore model on graph G with fugacity $\lambda \leq (1 - \delta)\lambda_c(\Delta_G)$, as discussed in Example 1.10, the Gibbs distribution μ is completely $O(\frac{1}{\delta})$-spectrally independent. We choose $\theta = \frac{1}{25}$. Due to Lemma 2.3, the field dynamics P_θ^{FD} for μ has a spectral gap $1/C_\delta$ where $C_\delta = \exp(O(\frac{1}{\delta}))$, and thus mixes in $\tilde{O}(C_\delta \cdot n)$ steps. Meanwhile, each transition step of P_θ^{FD} is simulated as a subroutine by a Glauber dynamics for $\pi = \mu^{(\theta)}$ that corresponds to a hardcore model with fugacity $\theta\lambda = \lambda/25 < \frac{1}{2\Delta_G}$, which mixes in $\tilde{O}(n)$ steps due to path coupling. Together this gives us a sampling algorithm for the hardcore model that runs in $\tilde{O}(C_\delta \cdot n^2)$ steps of single-site updates regardless of the maximum degree as long as $\lambda \leq (1 - \delta)\lambda_c(\Delta_G)$.*

For general anti-ferromagnetic two-spin systems, similar results hold as discussed in Remark 1.11.

This sampling procedure is different from the Glauber dynamics for μ. Nevertheless, the rapid mixing of this new procedure might serve as a "proof of concept" for the rapid mixing of the Glauber dynamics. Indeed, compared to the Glauber dynamics, in this new procedure, the $(+1)$-variables are less favored than (-1)-variables when being chosen to be resampled; but when being resampled, variables have higher chances to be updated to -1 than in the Glauber dynamics for μ, because they are now being resampled according to $\pi = \mu^{(\theta)}$ for $\theta \in (0, 1)$. These two types of biases might be canceling each other. Conceptually, a rapidly mixing Glauber dynamics is somehow conceived in this thought experiment.

Such intuition is formally justified by a comparison of variance decays of the chains. Given any function $f \in \mathbb{R}^{\Omega(\mu)}$, let $\text{Var}_\mu[f]$ and $\mathcal{E}_P(f, f)$ respectively denote its variance and Dirichlet form (defined in Section 3.3). It is well known that $\lambda_{\text{gap}}(P) = \inf_f \frac{\mathcal{E}_P(f,f)}{\text{Var}_\mu[f]}$ for any reversible chain P.

By comparing the Dirichlet forms of the field dynamics P_θ^{FD} and the Glauber dynamics P_μ, we establishes

$$(5) \qquad \mathcal{E}_{P_\theta^{\text{FD}}}(f, f) \leq \frac{1}{\lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(\theta)}\right)} \mathcal{E}_{P_\mu}(f, f).$$

This is proved in Section 4, although there it is expressed as a tensorization of variance for μ, which is equivalent to the comparison of Dirichlet forms. The comparison lemma for the field dynamics (Lemma 2.4) follows directly from (5) and characterization of spectral gap $\lambda_{\text{gap}}(P) = \inf_f \frac{\mathcal{E}_P(f,f)}{\text{Var}_\mu[f]}$.

2.4. Wrapping up.

We now describe how to prove the main theorem for anti-ferromagnetic two-spin systems (Theorem 1.1) by the main technical theorem (Theorem 1.9).

In general, magnetizing a distribution μ over $\{-1, +1\}^V$ to $\mu^{(\theta)}$ for a $\theta \in (0, 1)$ does not necessarily make it easier for sampling. Nevertheless, for anti-ferromagnetic two-spin systems, this issue can be circumvented by flipping roles of $+1$ and -1 for certain variables. Given a *direction vector* $\chi \in \{-1, +1\}^V$, the distribution $\nu = \text{flip}(\mu, \chi)$ obtained from *flipping* μ according to χ is defined as:

$$\forall \sigma \in \{-1, +1\}^V, \quad \nu(\sigma) \triangleq \mu(\sigma \odot \chi),$$

where $\sigma \odot \chi \in \{-1, +1\}^V$ is defined as that $(\sigma \odot \chi)_v = \sigma_v \chi_v$ for all $v \in V$.

Now consider an anti-ferromagnetic two-spin system on graph $G = (V, E)$, specified by parameters (β, γ, λ), where $0 \leq \beta \leq \gamma$, $\beta\gamma < 1$, and $\gamma, \lambda > 0$. The good direction $\chi \in \{-1, +1\}^V$ is constructed as:

$$(6) \qquad \forall v \in V, \quad \chi_v \triangleq \begin{cases} +1 & \lambda \leq \left(\frac{\gamma}{\beta}\right)^{\Delta_v/2} \\ -1 & \text{otherwise,} \end{cases}$$

where Δ_v denotes the degree of vertex v.

Assuming that the anti-ferromagnetic two-spin system is up-to-Δ unique with gap δ (Definition 3.2), the following theorem guarantees that after flipping according to the good direction χ in (6), the Gibbs distribution μ becomes completely $O(\frac{1}{\delta})$-spectrally independent, and if further every local field is biased by a $\Theta(\delta^2)$-factor, the Glauber dynamics is known to be rapidly mixing with spectral gap $\Omega(\frac{\delta}{n})$.

Theorem 2.10. *For all $\delta \in (0, 1)$, for every anti-ferromagnetic two-spin system on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ that is up-to-Δ unique with gap δ, the distribution $\nu = \text{flip}(\mu, \chi)$ obtained from flipping the Gibbs distribution μ according to the direction χ defined in (6), satisfies:*

- *ν is completely $\frac{144}{\delta}$-spectrally independent;*
- *for $\theta = \frac{\delta^2}{64}$ and $C = \frac{8}{\delta}$, it holds that $\lambda_{\text{min-gap}}^{\text{GD}}(\pi) \geq \frac{1}{Cn}$, where $\pi = \nu^{(\theta)}$.*

Theorem 1.1 is an easy consequence of Theorem 2.10.

Proof of Theorem 1.1. By Theorem 2.10, we can apply Theorem 1.9 to distribution $\nu = \text{flip}(\mu, \chi)$, so that

$$\lambda_{\text{gap}}^{\text{GD}}(\nu) \geq \left(\frac{\delta^2}{128}\right)^{\frac{288}{\delta}+7} \cdot \lambda_{\text{min-gap}}^{\text{GD}}\left(\nu^{(\theta)}\right) \geq \left(\frac{\delta}{8\sqrt{2}}\right)^{\frac{590}{\delta}} \cdot \left(\frac{\delta}{8n}\right) = \frac{1}{C(\delta) \cdot n}, \quad \text{where } C(\delta) = \left(\frac{1}{\delta}\right)^{O(\frac{1}{\delta})}.$$

Note that distribution ν is isomorphic to the original Gibbs distribution μ and the transition matrices of the Glauber dynamics for μ and ν are equivalent up to a bijection, and thus have the same set of eigenvalues. Hence,

$$\lambda_{\text{gap}}^{\text{GD}}(\mu) = \lambda_{\text{gap}}^{\text{GD}}(\nu) \geq \frac{1}{C(\delta) \cdot n}, \quad \text{where } C(\delta) = \left(\frac{1}{\delta}\right)^{O(\frac{1}{\delta})}. \qquad \square$$

Theorem 2.10 is proved in Section 8 as follows:

- The first part regarding the complete spectral independence for the flipped distribution ν is proved by observing that the analyses in [?] and [?] remain to hold when each local field λ_v is further biased towards the direction indicated by the χ_v in (6).
- The second part regarding the spectral gap for the easier distribution $\pi = \nu^{(\theta)}$ is due to the decay of path coupling in that easier regime and its implication to the spectral gap [? , ch. 13.1].

More specifically, there is a good direction for the local fields such that: (1) the uniqueness, spatial mixing, and spectral independence properties are all closed in that direction; and (2) a natural relaxation of the uniqueness condition arises from standard path coupling once the system moves towards the good direction by a constant factor. These discoveries may suggest that the notion of "easier regime" is perhaps more naturally described in our fashion, as biasing the local fields towards its easier direction.

The mixing time bound in (2) follows from (1) and the following straightforward analysis of $\mu_{\min}$.

Proof of (2). Define the marginal lower bound by

$$b \triangleq \min_{\Lambda \subseteq V, v \in V \setminus \Lambda} \min_{\sigma \in \Omega(\mu_\Lambda)} \min_{c \in \Omega(\mu_v^\sigma)} \mu_v^\sigma(c).$$

Obviously $\mu_{\min} \geq b^n$ due to chain rule, hence $\log \frac{1}{\mu_{\min}} \leq n \log \frac{1}{b}$.

When $\beta = 0$, the marginal bound has $b \geq \min_{1 \leq d \leq \Delta} \left\{ \frac{\gamma^d}{\gamma^d + \lambda}, \frac{\lambda}{\gamma^d + \lambda} \right\}$, which implies $\frac{1}{b} \leq (\lambda + \frac{1}{\lambda})\left(\frac{1}{\gamma} + \gamma + 2\right)^\Delta$.

When $\beta > 0$, by considering the worst configuration of the neighborhood, we have

$$b \geq \min_{1 \leq d \leq \Delta} \min_{0 \leq s \leq d} \min \left\{ \frac{\gamma^s}{\gamma^s + \lambda\beta^{d-s}}, \frac{\lambda\beta^{d-s}}{\gamma^s + \lambda\beta^{d-s}} \right\},$$

which implies that

$$\frac{1}{b} \leq \begin{cases} 2 + \lambda + \frac{1}{\lambda}\beta^{-\Delta} & \leq (\frac{1}{\lambda} + \lambda)\left(\frac{1}{\beta} + 2\right)^\Delta & \text{if } \gamma \geq 1, \\ 2 + \lambda\gamma^{-\Delta} + \frac{1}{\lambda}\beta^{-\Delta} & \leq (\frac{1}{\lambda} + \lambda)\left(\frac{1}{\beta} + 2\right)^\Delta & \text{if } \gamma < 1. \end{cases}$$

Overall, $\log \frac{1}{\mu_{\min}} \leq n \log \left(\lambda + \frac{1}{\lambda}\right) + n\Delta \log \alpha$, where

$$\alpha = \begin{cases} \gamma + \frac{1}{\gamma} + 2 & \text{if } \beta = 0, \\ \frac{1}{\beta} + 2 & \text{if } \beta > 0. \end{cases}$$

The mixing time bound in (2) follows from (1). $\square$

For the hardcore mode and the Ising model, similar results hold with better choice of θ and tighter bounds of $\mu_{\min}$, which are proved in Section 8.3. Hence we have Theorem 1.3 and Theorem 1.4.

2.5. **Related work and open problems**. Complexity classification of two-spin systems is a fundamental problem [? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?]. Previously, the primary technique for relating the spatial mixing properties to the rapid mixing of Glauber dynamics for spin systems is coupling [? ? ? ? ?]. In [?], techniques based on high-dimensional expanders [? ? ? ?] were applied to bound the mixing time of Glauber dynamics for 2-spin systems through the notion of spectral independence. The results were substantially improved and strengthened to an optimal $O(n \log n)$ mixing time bound for $\Delta = O(1)$ in [? ?] through entropy factorization [? ?]. These were extended to general spin models beyond Boolean domain and/or more general class of Markov chains [? ? ? ? ?].

In this paper, we prove an optimal $\Omega(n^{-1})$ lower bound on the spectral gap of the Glauber dynamics for anti-ferromagnetic two-spin systems satisfying the uniqueness condition with a constant slack. It leaves several open directions. First, a major open problem is to prove an optimal $O(n \log n)$ mixing time bound for the same regime with no degree restriction. A powerful tool for this goal is a "local-to-global" argument for relative entropy decays (modified log-Sobolev constants) as given in [? ?]. However, a major obstacle for this approach is that the current analyses of the uniform block dynamics based on entropy decay result in bounds that grow exponentially in the max-degree Δ. And even if this is resolved, another difficulty is to establish modified log-Sobolev inequalities in the "easier" regime.

Our techniques crucially rely on variables with Boolean domain. It is important to extend our approach to general distributions with variables beyond Boolean domains, e.g. proper q-colorings.

Finally, an open aspect is to optimize the reliance on δ. In fact, it should be restated as to improve the reliance of the mixing time on the spectral independence. To see this is an important question, consider graph matchings, whose spectral independence is bounded by $O(\sqrt{\Delta})$, but so far in this entire line of research the dependency of mixing time on spectral independence is at least exponential.

2.6. **Organization of the paper**. The preliminaries are given in Section 3. The comparison lemma for the field dynamics (Lemma 2.4) is proved in Section 4. And the mixing lemma for the field dynamics (Lemma 2.3) is proved in Section 5 (for preparation), Section 6 (for the proof of Lemma 2.7) and Section 7 (for the proof of Lemma 2.8). Finally, the mixing results for the two-spin systems are proved in Section 8. Additionally, In Appendix A, we provide a proof of Theorem 2.5 for the mixing of uniform block dynamics assuming our notion of spectral independence with absolute influence matrix; and more proofs for the uniqueness and spectral independence of two-spin systems with local fields are provided in Appendix B.

3. Preliminaries

3.1. **Notations**. Throughout the paper, we use log to denote the natural logarithm with base e.

Let V be a ground set. For any configuration $\sigma \in \{-1, +1\}^V$, we use $\sigma^{-1}(c) \triangleq \{v \in V \mid \sigma_v = c\}$ to denote the pre-image of $c \in \{-1, +1\}$ under σ. Let $\|\sigma\|_+ \triangleq |\sigma^{-1}(+1)|$ (and $\|\sigma\|_- \triangleq |\sigma^{-1}(-1)|$) denote the number of $+1$'s (and -1's) in σ. For $\Lambda \subseteq V$, let $\mathbf{1}_\Lambda \in \{+1\}^\Lambda$ denote the all-$(+1)$ configuration on Λ.

For a probability distribution μ, we use $\Omega(\mu)$ to denote the support of μ.

Let $G = (V, E)$ be an undirected graph. For every vertex $v \in V$, we use Δ_v to denote the degree of v in G and denote by $\Delta_G \triangleq \max_{v \in V} \Delta_v$ the maximum degree of G.

3.2. **The uniqueness condition**. Let β, γ, λ be real numbers satisfying

$$(7) \qquad\qquad 0 \le \beta \le \gamma, \gamma > 0, \lambda > 0 \text{ and } \beta\gamma < 1,$$

that is, (β, γ, λ) gives parameters for an anti-ferromagnetic two-spin system.

Given any integer $d \ge 1$, the univariate tree recursion $F_d : \mathbb{R} \to \mathbb{R}$ is defined by

$$(8) \qquad\qquad F_d(x) \triangleq \lambda \left(\frac{\beta x + 1}{x + \gamma} \right)^d,$$

and let $\hat{x}_d$ denote the unique positive fixed point of F_d, i.e. $\hat{x}_d = F_d(\hat{x}_d)$.

Definition 3.1 (d-uniqueness [?]). Let $\delta \in (0, 1)$, and $d \ge 1$ be an integer. A (β, γ, λ), where β, γ, λ satisfy (7), is said to be *d-unique with gap δ* if

$$(9) \qquad\qquad f_d(\hat{x}_d) \triangleq |F_d'(\hat{x}_d)| = \frac{d(1 - \beta\gamma)\hat{x}_d}{(\beta\hat{x}_d + 1)(\hat{x}_d + \gamma)} \le 1 - \delta.$$

Definition 3.2 (up-to-Δ uniqueness [?]). Let $\delta \in (0, 1)$ and $\Delta \in [3, +\infty]$. A (β, γ, λ), where β, γ, λ satisfy (7), is said to be *up-to-Δ unique with gap δ* if it is d-unique with gap δ for all integers $1 \le d < \Delta$.

Note that the definition includes that $\Delta = +\infty$, in which case the d-uniqueness should hold for all $d \ge 1$.

Let $\mathcal{I} = (V, E, \beta, \gamma, \lambda)$ be an anti-ferromagnetic two-spin system specified by parameters (β, γ, λ) satisfying (7), on graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \ge 3$. If (β, γ, λ) is up-to-Δ unique with gap δ, we simply say that $\mathcal{I}$ is up-to-Δ unique with gap δ.

3.3. **Markov chain, spectral gap and coupling**.

3.3.1. *Basic definitions*. Let Ω be a finite state space. Let $(X_t)_{t \ge 0}$ be a Markov chain over Ω with *transition matrix* $P \in \mathbb{R}_{\ge 0}^{\Omega \times \Omega}$. We use matrix P to refer to the corresponding Markov chain if this is clear in the context. The Markov chain is *irreducible* if for any $X, Y \in \Omega$, there is an integer t such that $P^t(X, Y) > 0$. The Markov chain is *aperiodic* if for any $X \in \Omega$, $\gcd\{t > 0 \mid P^t(X, X) > 0\} = 1$. A distribution μ is called a stationary distribution of P if $\mu = \mu P$. If a Markov chain is both irreducible and aperiodic, then it has a unique stationary distribution. The Markov chain P is *reversible* with respect to a distribution μ if the following *detailed balance equation* is satisfied

$$\forall X, Y \in \Omega, \quad \mu(X)P(X, Y) = \mu(Y)P(Y, X),$$

which implies μ is a stationary distribution of P.

Let μ be a distribution with support Ω. Let P be a Markov chain over Ω with the unique stationary distribution μ. The *mixing time* of P is defined by

$$\forall 0 < \varepsilon < 1, \quad T_{\mathrm{mix}}(\varepsilon) \triangleq \max_{X \in \Omega} \min \left\{ t \mid d_{\mathrm{TV}} \left(P^t(X, \cdot), \mu \right) \le \varepsilon \right\},$$

where $P^t(X, \cdot)$ is the distribution generated by the Markov chain after t transition steps when starting from X, and $d_{\mathrm{TV}} \left(P^t(X, \cdot), \mu \right)$ denotes the *total variation distance* between $P^t(X, \cdot)$ and μ, formally,

$$d_{\mathrm{TV}} \left(P^t(X, \cdot), \mu \right) \triangleq \frac{1}{2} \sum_{Y \in \Omega} \left| P^t(X, Y) - \mu(Y) \right|.$$

3.3.2. *Spectral gap of reversible Markov chains.* For reversible Markov chains, the mixing time is closely related to the *spectral gap*. Let μ be a distribution with support $\Omega(\mu)$. For any $f, g \in \mathbb{R}^{\Omega(\mu)}$, define their *inner product* with respect to μ by

$$\langle f, g \rangle_\mu \triangleq \sum_{\sigma \in \Omega(\mu)} \mu(\sigma) f(\sigma) g(\sigma).$$

Let P be a Markov chain over $\Omega = \Omega(\mu)$ that is reversible with respect to μ. It is well known that the transition matrix P is a self-adjoint matrix with respect to inner product $\langle \cdot, \cdot \rangle_\mu$, formally,

$$\forall f, g \in \mathbb{R}^{\Omega(\mu)}, \quad \langle Pf, g \rangle_\mu = \langle f, Pg \rangle_\mu.$$

By standard linear algebra results [? , Lemma 12.2], P has $|\Omega|$ real eigenvalues $1 = \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{|\Omega|} \ge -1$; each λ_i corresponds to a real eigenvector $f_i \in \Omega^{\mathbb{R}}$ such that $P f_i = \lambda_i f_i$, where $f_1 = \mathbf{1}$ and $f_1, f_2, \ldots, f_{|\Omega|}$ form orthonormal bases of inner product space $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\mu)$. Assume $|\Omega| \ge 2$. The *absolute spectral gap* of P is defined by $1 - \lambda_\star \triangleq 1 - \max\{|\lambda_i| \mid 2 \le i \le |\Omega|\}$. The *spectral gap* $\lambda_{\mathrm{gap}}(P)$ of P is defined by $1 - \lambda_2$. We simply assume $\lambda_\star = \lambda_2 = 0$ if $|\Omega| = 1$. The following relation between mixing time and absolute spectral gap is well-known [? , Theorem 12.4]:

$$(10) \qquad T_{\mathrm{mix}}(\varepsilon) \le \frac{1}{1 - \lambda_\star} \log \left(\frac{1}{\varepsilon \mu_{\mathrm{min}}} \right), \quad \text{where } \mu_{\mathrm{min}} = \min_{\sigma \in \Omega(\mu)} \mu(\sigma).$$

To analyze the spectral gap of a reversible Markov chain P, we introduce the following standard notations. Let $f \in \mathbb{R}^{\Omega(\mu)}$. The 2-norm of function $f \in \mathbb{R}^{\Omega(\mu)}$ with respect to μ is defined by

$$\|f\|_{2,\mu} \triangleq \sqrt{\langle f, f \rangle_\mu},$$

the *expectation* of f with respect to μ is defined by

$$(11) \qquad \mathbf{E}_\mu [f] \triangleq \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma),$$

the *variance* with respect to μ is defined by

$$(12) \qquad \mathbf{Var}_\mu [f] \triangleq \mathbf{E}_\mu \left[f^2 \right] - (\mathbf{E}_\mu [f])^2 = \frac{1}{2} \sum_{\sigma, \tau \in \Omega} \mu(\sigma) \mu(\tau) \left(f(\sigma) - f(\tau) \right)^2,$$

and the *Dirichlet form* with respect to P (where P is reversible with respect to μ) is defined by

$$(13) \qquad \mathcal{E}_P(f, f) \triangleq \langle f, (I - P)f \rangle_\mu = \frac{1}{2} \sum_{\sigma, \tau \in \Omega} \mu(\sigma) P(\sigma, \tau) \left(f(\sigma) - f(\tau) \right)^2.$$

We can slightly extend above definitions to allow functions $f \in \mathbb{R}^S$ for some $S \supseteq \Omega$, for these functions f, the expectation in (11), the variance in (12) and the Dirichlet form in (13) are all well-defined.

For any $f, g \in \mathbb{R}^\Omega$, we write $f \perp_\mu g$ if $\langle f, g \rangle_\mu = 0$. By Courant-Fischer theorem [? , Lemma 13.7],

$$\lambda_{\mathrm{gap}}(P) = \inf \left\{ \mathcal{E}_P(f, f) \mid f \in \mathbb{R}^{\Omega(\mu)}, f \perp_\mu \mathbf{1}, \|f\|_{2,\mu} = 1 \right\}.$$

Furthermore, the spectral gap can also be characterized as [? , Remark 13.8]:

$$(14) \qquad \lambda_{\text{gap}}(P) = \inf \left\{ \frac{\mathcal{E}_P(f, f)}{\text{Var}_\mu [f]} \mid f \in \mathbb{R}^{\Omega(\mu)}, \text{Var}_\mu [f] \neq 0 \right\},$$

The Poincaré inequality follows from the above equation:

$$(15) \qquad \forall f \in \mathbb{R}^{\Omega(\mu)}, \quad \mathcal{E}_P(f, f) \geq \lambda_{\text{gap}}(P) \cdot \text{Var}_\mu [f].$$

3.3.3. *Coupling of Markov chains.* Let μ and ν be two distributions on Ω. A coupling of μ and ν is a joint distribution (X, Y) over $\Omega \times \Omega$ such that the marginal distributions of X and Y are μ and ν respectively. The following is the well-known coupling lemma.

Lemma 3.3 ([? , Proposition 4.7]). *For any coupling (X, Y) of μ and ν,*

$$\mathbf{Pr}[X \neq Y] \geq d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)|.$$

Furthermore, there is an optimal coupling (X, Y) such that $\mathbf{Pr}[X \neq Y] = d_{\text{TV}}(\mu, \nu)$.

Let P denote a Markov chain over the state space Ω. A coupling of Markov chain is a joint stochastic process $(X_t, Y_t)_{t \geq 0}$ such that each individual process $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ follow the transition rule of P, and if $X_t = Y_t$, then $X_k = Y_k$ for all $k \geq t$. The following lemma connects the spectral gap of Markov chain and the contraction rate in coupling.

Lemma 3.4 ([?] and [? , ch. 13.1]). *Let μ be a distribution with support Ω, and Φ a metric on Ω, where $|\Omega| \geq 2$. Let $P \in \mathbb{R}_{\geq 0}^{\Omega \times \Omega}$ be the transition matrix of a Markov chain that is reversible with respect to μ. If there exists $0 < r < 1$ such that for any $X, Y \in \Omega$, there exists a coupling $(X, Y) \to (X', Y')$ of P such that*

$$\mathbf{E}[\Phi(X', Y') \mid X, Y] \leq (1 - r)\Phi(X, Y),$$

then the spectral gap of P satisfies

$$1 - \lambda_2(P) \geq r,$$

where $\lambda_2(P)$ is the second largest eigenvalue of P.

3.4. Uniform block dynamics and Glauber dynamics.
Let μ be a distribution over $\{-1, +1\}^V$, where V is a ground set. For any integer $1 \leq \ell \leq |V|$, let $\binom{V}{\ell} \triangleq \{S \subseteq V \mid |S| = \ell\}$ denote the collection of all subsets of size ℓ. The (heat-bath) uniform block dynamics for μ is defined as follows.

Definition 3.5 (Heat-bath uniform ℓ-block dynamics). For any positive integer $1 \leq \ell \leq |V|$, the (heat-bath) uniform ℓ-block dynamics for μ is a Markov chain $(X_t)_{t \geq 0}$ over $\Omega(\mu)$. The chain starts from an arbitrary configuration $X_0 \in \Omega$. In the t-th transition step, the chain evolves as follows:
- pick a set $S \in \binom{V}{\ell}$ uniformly at random, and set $X_t(V \setminus S) = X_{t-1}(V \setminus S)$;
- sample $X_t \sim \mu_S^{X_t(V \setminus S)}$, where $\mu_S^{X_t(V \setminus S)}$ denotes the marginal distribution on S induced from μ conditional on the assignment of $X_t(V \setminus S)$.

In particular, the uniform 1-block-dynamics is known as the *Glauber dynamics* for μ.

The following proposition was known.

Proposition 3.6 ([? ? ?]). *Let μ be a distribution over $\{-1, +1\}^V$. For any integer $1 \leq \ell \leq |V|$, the uniform ℓ-block dynamics for μ is reversible with respect to μ, and its transition matrix is positive semidefinite.*

Let μ be a distribution over $\{-1, +1\}^V$ and $1 \leq \ell \leq |V|$ an integer. Let P denote the transition matrix of the uniform ℓ-block dynamics for μ.

$$T_{\text{mix}}(\varepsilon) \leq \frac{1}{1 - \lambda_2(P)} \log \left(\frac{1}{\varepsilon \mu_{\min}} \right), \quad \text{where } \mu_{\min} = \min_{\sigma \in \Omega(\mu)} \mu(\sigma),$$

where $\lambda_2(P)$ is the second largest eigenvalue of P.

For Glauber dynamics (uniform 1-block dynamics), its spectral gap can be further characterized by the *approximate tensorization of variance*. For any variable $v \in V$, and any function $f \in \mathbb{R}^{\Omega(\mu)}$, define

$$\mu[\mathbf{Var}_v[f]] \triangleq \sum_{\sigma \in \Omega(\mu_{V \setminus \{v\}})} \mu_{V \setminus \{v\}}(\sigma) \mathbf{Var}_{\mu^\sigma}[f],$$

where μ^σ is the distribution over $\{-1, +1\}^V$ obtained from μ conditional on the configuration on $V \setminus \{v\}$ being fixed as σ. We remark that $\mu[\mathbf{Var}_v[f]]$ is also well-defined if $f \in \mathbb{R}^S$ for some $S \supseteq \Omega(\mu)$. The approximate tensorization of variance is defined as follows.

Definition 3.7. Let $C > 0$ be a parameter. A distribution μ over $\{-1, +1\}^V$ with $|\Omega(\mu)| \geq 2$ is said to satisfy the *approximate tensorization of variance* with parameter C if for all $f \in \mathbb{R}^{\Omega(\mu)}$,

$$\mathbf{Var}_\mu[f] \leq C \sum_{v \in V} \mu[\mathbf{Var}_v[f]].$$

Lemma 3.8 (Fact A.3 [?]). *A distribution μ over $\{-1, +1\}^V$ with $|\Omega(\mu)| \geq 2$ satisfies the approximate tensorization of variance with parameter C if and only if the spectral gap $\lambda_{\mathrm{gap}}(P) \geq \frac{1}{Cn}$, where $n = |V|$, and P denotes the Glauber dynamics for μ.*

3.5. **Multivariate hypergeometric distribution.** Let V be a set of n buckets, each of them has k balls. Suppose we pick ℓ balls from all kn balls uniformly at random, without replacement. For each bucket $v \in V$, let $a_v \in \mathbb{Z}_{\geq 0}$ denote the number of balls picked from the bucket v, then $\boldsymbol{a} = (a_v)_{v \in V}$ follows multivariate hypergeometric distribution.

Formally, given a set V of size n, an integer $k \geq 1$ and an integer $0 \leq \ell \leq kn$, the multivariate hypergeometric distribution $\Pi_{V,k,\ell}$ is defined as follows. The support of $\Pi_{V,k,\ell}$ is defined by

$$(16) \qquad \Omega(\Pi_{V,k,\ell}) \triangleq \left\{ \boldsymbol{a} = (a_v)_{v \in V} \mid \sum_{v \in V} a_v = \ell \text{ and } \forall v \in V, a_v \in \mathbb{Z}_{\geq 0} \right\},$$

For any $\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell})$, it holds that

$$(17) \qquad \Pi_{V,k,\ell}(\boldsymbol{a}) = \frac{\prod_{v \in V} \binom{k}{a_v}}{\binom{kn}{\ell}}.$$

For any $i \in [kn] \triangleq \{1, 2, 3, \ldots, kn\}$, we use X_i to indicate whether the i-th ball is picked. The random variables $X_1, X_2, \cdots, X_{kn}$ are negatively associated [?], and hence the Chernoff-Hoeffding inequality is applicable.

Lemma 3.9 ([?] and [?]). *Let $n \geq 1, k \geq 1$ be integers. Given an integer $1 \leq \ell \leq kn$, let $\mathcal{S} \in \binom{[kn]}{\ell}$ be a random subset sampled from $\binom{[kn]}{\ell}$ uniformly at random.*

For each $1 \leq i \leq kn$, define the indicator random variable $X_i = \mathbf{1}[i \in \mathcal{S}]$. Let $X = \sum_{i=1}^{k} X_i$. For any $\varepsilon \in (0, 1)$ it holds that

$$\mathbf{Pr}\left[\left| \frac{X}{k} - \frac{\ell}{kn} \right| \geq \varepsilon \right] \leq 2 \exp\left(-2\varepsilon^2 k\right).$$

Proof. Since $X_1, X_2, \ldots, X_k$ are negative associated [? , Lemma 2.11], the Chernoff-Hoeffding inequality [?] can be applied to their summation X. $\square$

4. Comparing Glauber Dynamics with Field Dynamics

In this section, we prove the comparison lemma for the field dynamics (Lemma 2.4).

Let μ be a distribution over $\{-1, +1\}^V$ and $\theta \in (0, 1)$. Let $\pi = \mu^{(\theta)}$, which is defined as in Definition 1.7. Note that μ and π have the same support $\Omega \triangleq \Omega(\pi) = \Omega(\mu)$ for $\theta > 0$. Formally, π is defined as:

$$
(18) \qquad \forall \sigma \in \{-1, +1\}^V, \quad \pi(\sigma) = \frac{\mu(\sigma)\theta^{\|\sigma\|_+}}{Z_\pi}, \quad \text{where } Z_\pi \triangleq \sum_{\sigma \in \{-1, +1\}^V} \mu(\sigma)\theta^{\|\sigma\|_+}.
$$

Moreover, we use $\mathbf{Pr}_{R \subseteq V}[\cdot]$ to denote the law for subset $R \subseteq V$ that is randomly generated by including each $v \in V$ into R independently with probability $1 - \theta$. Specifically, for every $\Lambda \subseteq V$,

$$
(19) \qquad \Pr_{R \subseteq V}[R = \Lambda] = (1 - \theta)^{|\Lambda|}\theta^{|V| - |\Lambda|}.
$$

Let P_θ^{FD} denote the field dynamics for distribution μ with parameter θ. The Dirichlet form of P_θ^{FD} can be calculated as follows.

Lemma 4.1. *Let $\pi = \mu^{(\theta)}$ be defined as in* (18)*. For all $f \in \mathbb{R}^\Omega$,*

$$
(20) \qquad \mathcal{E}_{P_\theta^{\mathrm{FD}}}(f, f) = \frac{Z_\pi}{\theta^{|V|}} \mathop{\mathbf{E}}_{R \subseteq V}\left[\pi_R(\mathbf{1}_R) \cdot \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]\right].
$$

where $\mathbf{1}_R \in \{-1, +1\}^R$ denotes the all-$(+1)$ configuration specified on $R \subseteq V$, and the expectation is calculated assuming the convention that $\pi_R(\mathbf{1}_R)\mathbf{Var}_{\pi^{\mathbf{1}_R}}[f] = 0$ when $\pi_R(\mathbf{1}_R) = 0$.[1]

Intuitively, the $R \subseteq V$ corresponds to $R = V \setminus S$ where S is the random set of resampled variables generated in the transition step of the field dynamics, such that every $v \in V$ with current value -1 is selected into S, and thus the configuration over $R = V \setminus S$ must be $\mathbf{1}_R$. And the quantity $\mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]$ arises because the current configuration over S is resampled according to $\pi = \mu^{(\theta)}$ conditional on the current configuration over $R = V \setminus S$, which is $\mathbf{1}_R$. The proof of Lemma 4.1 is postponed to the end of this section.

Consider the Glauber dynamics for each $\pi^{\mathbf{1}_R}$. By Lemma 3.8, for all $R \subseteq V$ with $\pi_R(\mathbf{1}_R) > 0$, the distribution $\pi^{\mathbf{1}_R}$ satisfies the approximate tensorization of variance with parameter $\frac{1}{n \cdot \lambda_{\mathrm{gap}}^{\mathrm{GD}}(\pi^{\mathbf{1}_R})}$:

$$
\forall f \in \mathbb{R}^\Omega, \quad \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f] \le \frac{1}{n \cdot \lambda_{\mathrm{gap}}^{\mathrm{GD}}(\pi^{\mathbf{1}_R})} \sum_{v \in V} \pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]] \le \frac{1}{n \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\pi)} \sum_{v \in V} \pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]],
$$

where $\lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\pi) \le \lambda_{\mathrm{gap}}^{\mathrm{GD}}(\pi^\sigma)$ for all feasible partial configuration σ, as defined in (4).

Then by (20), the Dirichlet form $\mathcal{E}_{P_\theta^{\mathrm{FD}}}(f, f)$ is upper bounded as

$$
(21) \qquad \mathcal{E}_{P_\theta^{\mathrm{FD}}}(f, f) \le \frac{Z_\pi}{n \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\pi) \cdot \theta^{|V|}} \sum_{v \in V} \mathop{\mathbf{E}}_{R \subseteq V}\left[\pi_R(\mathbf{1}_R) \cdot \pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]]\right].
$$

Let $I[\cdot]$ denote the indicator variable. The following identity holds for all $f \in \mathbb{R}^\Omega$:

$$
\mathop{\mathbf{E}}_{R \subseteq V}\left[\pi_R(\mathbf{1}_R) \cdot \pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]]\right] \overset{(\star)}{=} \mathop{\mathbf{E}}_{R \subseteq V}\left[\mathop{\mathbf{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[I\left[R \subseteq \sigma^{-1}(+1)\right] \cdot \mathbf{Var}_{\pi^\sigma}[f]\right]\right]
$$

$$
= \mathop{\mathbf{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[\Pr_{R \subseteq V}\left[R \subseteq \sigma^{-1}(+1)\right] \cdot \mathbf{Var}_{\pi^\sigma}[f]\right]
$$

$$
(22) \qquad = \mathop{\mathbf{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[\theta^{|V| - \|\sigma\|_+} \cdot \mathbf{Var}_{\pi^\sigma}[f]\right],
$$

where the nontrivial equation $(\star)$ holds by verifying for every choice of $v \in V$ and $R \subseteq V$ as follows:

---

[1]Equivalently, one may think this as an expectation $\mathbf{E}_{R \subseteq V}[\mathcal{P}(R)]$ of a function $\mathcal{P}(R)$, where $\mathcal{P}(R) = 0$ if $\pi_R(\mathbf{1}_R) = 0$ and $\mathcal{P}(R) = \pi_R(\mathbf{1}_R)\mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]$ if $\pi_R(\mathbf{1}_R) > 0$, in which case the variance $\mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]$ is well-defined.

- For the case that $\pi_R(\mathbf{1}_R) > 0$ and $v \notin R$, it holds that

$$\pi_R(\mathbf{1}_R) = \Pr_{\sigma \sim \pi}\left[R \subseteq \sigma^{-1}(+1)\right] = \Pr_{\sigma \sim \pi_{V \setminus \{v\}}}\left[R \subseteq \sigma^{-1}(+1)\right],$$

and the variance $\pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]]$ is well-defined, such that

$$\pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]] = \mathop{\mathbb{E}}_{\sigma \sim \pi^{\mathbf{1}_R}_{V \setminus \{v\}}}[\mathbf{Var}_{\pi^\sigma}[f]] = \mathop{\mathbb{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[\mathbf{Var}_{\pi^\sigma}[f] \mid R \subseteq \sigma^{-1}(+1)\right].$$

Therefore,

$$\pi_R(\mathbf{1}_R) \cdot \pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]] = \Pr_{\sigma \sim \pi_{V \setminus \{v\}}}\left[R \subseteq \sigma^{-1}(+1)\right] \cdot \mathop{\mathbb{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[\mathbf{Var}_{\pi^\sigma}[f] \mid R \subseteq \sigma^{-1}(+1)\right]$$

$$= \mathop{\mathbb{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[I\left[R \subseteq \sigma^{-1}(+1)\right] \cdot \mathbf{Var}_{\pi^\sigma}[f]\right].$$

- For the case that $\pi_R(\mathbf{1}_R) = 0$ or $v \in R$, both sides are 0. On the left-hand-side, if $\pi_R(\mathbf{1}_R) = 0$, then by convention

$$\pi_R(\mathbf{1}_R) \cdot \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f] = 0;$$

or else, if $\pi_R(\mathbf{1}_R) > 0$ but $v \in R$, then the variance $\pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]]$ is well-defined, but for $\sigma \sim \pi^{\mathbf{1}_R}_{V \setminus \{v\}}$, the $\mathbf{1}_R \uplus \sigma$ gives a configuration fully specified on V and hence the variance becomes trivial, i.e.

$$\pi^{\mathbf{1}_R}[\mathbf{Var}_v[f]] = \mathop{\mathbb{E}}_{\sigma \sim \pi^{\mathbf{1}_R}_{V \setminus \{v\}}}\left[\mathbf{Var}_{\pi^{\mathbf{1}_R \uplus \sigma}}[f]\right] = 0.$$

On the right-hand-side, if $\pi_R(\mathbf{1}_R) = 0$ or $v \in R$, then for $\sigma \sim \pi_{V \setminus \{v\}}$, the event $R \subseteq \sigma^{-1}(+1)$ can never occur, and hence

$$\mathop{\mathbb{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[I\left[R \subseteq \sigma^{-1}(+1)\right] \cdot \mathbf{Var}_{\pi^\sigma}[f]\right] = 0.$$

This gives the equation $(\star)$ in (22). Meanwhile, the other two equations in (22) follows respectively from linearity of expectation and the fact that $\Pr_{R \subseteq V}[R \subseteq \Lambda] = \theta^{|V| - |\Lambda|}$ for all $\Lambda \subseteq V$.

Furthermore, it can be verified that

$$\mathop{\mathbb{E}}_{\sigma \sim \pi_{V \setminus \{v\}}}\left[\frac{1}{\theta^{\|\sigma\|_+}} \mathbf{Var}_{\pi^\sigma}[f]\right] = \sum_{\sigma \in \Omega(\pi_{V \setminus \{v\}})} \frac{1}{\theta^{\|\sigma\|_+}} \pi_{V \setminus \{v\}}(\sigma) \pi^\sigma_v(-1) \pi^\sigma_v(+1)(f(\sigma_+) - f(\sigma_-))^2$$

$$= \frac{1}{Z_\pi} \sum_{\sigma \in \Omega(\pi_{V \setminus \{v\}})} \mu_{V \setminus \{v\}}(\sigma) \mu^\sigma_v(-1) \pi^\sigma_v(+1)(f(\sigma_+) - f(\sigma_-))^2$$

$$\leq \frac{1}{Z_\pi} \sum_{\sigma \in \Omega(\pi_{V \setminus \{v\}})} \mu_{V \setminus \{v\}}(\sigma) \mu^\sigma_v(-1) \mu^\sigma_v(+1)(f(\sigma_+) - f(\sigma_-))^2$$

(23)
$$= \frac{1}{Z_\pi} \mu[\mathbf{Var}_v[f]],$$

where $\sigma_\pm \in \{-1, +1\}^V$ denote the configurations on V where $\sigma_\pm(V \setminus \{v\}) = \sigma$ and $\sigma_\pm(v) = \pm 1$, the second equation is due to the chain rule and the relation between π and μ in (18), and the inequality is due to the relaxation

(24)
$$\pi^\sigma_v(+1) \leq \mu^\sigma_v(+1),$$

which holds because $\pi = \mu^{(\theta)}$ is obtained by biasing every variable with a local field $\theta \in (0, 1)$. Indeed,

$$\pi^\sigma_v(+1) = \frac{\pi(\sigma_+)}{\pi(\sigma_-) + \pi(\sigma_+)} = \frac{\theta \mu(\sigma_+)}{\mu(\sigma_-) + \theta \mu(\sigma_+)} \leq \frac{\mu(\sigma_+)}{\mu(\sigma_-) + \mu(\sigma_+)} = \mu^\sigma_v(+1),$$

where the inequality holds for $\theta \in (0, 1)$.

16

Combining (21), (22), and (23), we have the following upper bound on the Dirichlet form:

$$\mathcal{E}_{P_\theta^{\mathrm{FD}}}(f,f) \le \frac{1}{n \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\mu^{(\theta)})} \sum_{v \in V} \mu[\mathbf{Var}_v[f]].$$

Due to the Poincaré's inequality (15) for the field dynamics P_θ^{FD}, for all $f \in \mathbb{R}^\Omega$,

$$\mathbf{Var}_\mu[f] \le \frac{1}{\lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu,\theta)} \mathcal{E}_{P_\theta^{\mathrm{FD}}}(f,f) \le \frac{1}{n \cdot \lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu,\theta) \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\mu^{(\theta)})} \sum_{v \in V} \mu[\mathbf{Var}_v[f]].$$

This shows that μ satisfies the approximate tensorization of variance with parameter

$$\frac{1}{n \cdot \lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu,\theta) \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\mu^{(\theta)})}.$$

By Lemma 3.8, it gives us the following lower bound on the spectral gap of the Glauber dynamics for μ

$$\lambda_{\mathrm{gap}}^{\mathrm{GD}}(\mu) \ge \lambda_{\mathrm{gap}}^{\mathrm{FD}}(\mu,\theta) \cdot \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}\left(\mu^{(\theta)}\right),$$

which proves the comparison lemma for the field dynamics (Lemma 2.4).

Remark 4.2 (Tightness of Lemma 2.4). Apart from the Poincaré's inequalities, our proof of Lemma 2.4 is tight almost everywhere. The only exception is the relaxation of the marginal probability in (24).

Proof of Lemma 4.1. In each transition step of the field dynamics, a random subset $S \subseteq V$ of variables is generated for resampling. Let $R = V \setminus S$. The Dirichlet form can be calculated as

$$\mathcal{E}_{P_\theta^{\mathrm{FD}}}(f,f) = \frac{1}{2} \sum_{\sigma,\tau \in \Omega(\mu)} \mu(\sigma) P_\theta^{\mathrm{FD}}(\sigma,\tau)(f(\sigma) - f(\tau))^2$$

$$= \frac{1}{2} \sum_{\sigma,\tau \in \Omega(\mu)} \sum_{R \subseteq \sigma^{-1}(+1)} (1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \pi^{\mathbf{1}_R}(\tau) \mu(\sigma)(f(\sigma) - f(\tau))^2,$$

where the second equation is due to the following observations:

- R can only be a subset of $\sigma^{-1}(+1)$ since all variables $v \in V$ with $\sigma_v = -1$ are selected into $S = V \setminus R$;
- each $v \in \sigma^{-1}(+1)$ is independently selected into R with probability $1 - \theta$;
- given an $R \subseteq \sigma^{-1}(+1)$, the current configuration σ transits to τ with probability $\pi^{\sigma_R}(\tau) = \pi^{\mathbf{1}_R}(\tau)$.

Hence, fix any $R \subseteq V$, we only need to consider the pair $\sigma, \tau \in \Omega(\mu)$ such that $\sigma_R = \tau_R = \mathbf{1}_R$. We have

$$\mathcal{E}_{P_\theta^{\mathrm{FD}}}(f,f) = \frac{1}{2} \sum_{R \subseteq V: \mathbf{1}_R \in \Omega(\mu_R)} \left(\frac{1-\theta}{\theta}\right)^{|R|} \sum_{\sigma,\tau \in \Omega(\mu^{\mathbf{1}_R})} \theta^{\|\sigma\|_+} \pi^{\mathbf{1}_R}(\tau) \mu(\sigma)(f(\sigma) - f(\tau))^2.$$

By (18), it holds that $\pi(\sigma) Z_\pi = \mu(\sigma) \theta^{\|\sigma\|_+}$. We have

$$\mathcal{E}_{P_\theta^{\mathrm{FD}}}(f,f) = \frac{Z_\pi}{2} \sum_{R \subseteq V: \mathbf{1}_R \in \Omega(\pi_R)} \left(\frac{1-\theta}{\theta}\right)^{|R|} \sum_{\sigma,\tau \in \Omega(\pi^{\mathbf{1}_R})} \pi^{\mathbf{1}_R}(\tau) \pi(\sigma)(f(\sigma) - f(\tau))^2$$

$$= \frac{Z_\pi}{2} \sum_{R \subseteq V: \mathbf{1}_R \in \Omega(\pi_R)} \left(\frac{1-\theta}{\theta}\right)^{|R|} \pi_R(\mathbf{1}_R) \sum_{\sigma,\tau \in \Omega(\pi^{\mathbf{1}_R})} \pi^{\mathbf{1}_R}(\tau) \pi^{\mathbf{1}_R}(\sigma)(f(\sigma) - f(\tau))^2$$

$$= \frac{Z_\pi}{\theta^{|V|}} \sum_{R \subseteq V: \mathbf{1}_R \in \Omega(\pi_R)} (1-\theta)^{|R|} \theta^{|V|-|R|} \pi_R(\mathbf{1}_R) \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]$$

$$= \frac{Z_\pi}{\theta^{|V|}} \mathop{\mathbf{E}}_{R \subseteq V}\left[\pi_R(\mathbf{1}_R) \cdot \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f]\right],$$

where the second equation is due to the fact that $\sigma_R = \mathbf{1}_R$ for $\sigma \in \Omega(\pi^{\mathbf{1}_R})$ and the last equation adopts the convention that $\pi_R(\mathbf{1}_R) \mathbf{Var}_{\pi^{\mathbf{1}_R}}[f] = 0$ when $\pi_R(\mathbf{1}_R) = 0$. □

5. Mixing of Field Dynamics

In this section, and the next two sections, prove the mixing lemma for the field dynamics (Lemma 2.3). In Section 5.1, we prove Proposition 2.2 for the convergence and reversibility of the field dynamics. In Section 5.2, we start our proof of Lemma 2.7, which states that the field dynamics is the limiting instance for the uniform block dynamics on k-transformed distributions.

5.1. Convergence and reversibility (proof of Proposition 2.2). Fix a joint distribution μ over $\{-1, +1\}^V$. For any $0 < \theta < 1$, let P_θ^{FD} denote the transition matrix of the field dynamics for μ with parameter θ.

First, we verify that P_θ^{FD} is irreducible and aperiodic. Let $\pi = \mu^{(\theta)}$, which is as defined in Definition 1.7. Note that π and μ have the same support since $\theta > 0$, i.e.

$$\Omega \triangleq \Omega(\pi) = \Omega(\mu).$$

In each transition step, the chain constructs a random subset $S \subseteq V$ of variables that are going to be resampled. It holds that $S = V$ with positive probability for $\theta > 0$, in which case the current configuration σ is entirely resampled according to $\pi = \mu^{(\theta)}$. This means $P_\theta^{\mathsf{FD}}(\sigma, \tau) > 0$ for all pairs $\sigma, \tau \in \Omega$ of feasible configurations. Hence, the Markov chain is irreducible and aperiodic.

Next, we prove the reversibility of the chain with respect to μ by verifying the detailed balance equation:

$$(25) \qquad \forall \sigma, \tau \in \Omega, \quad \mu(\sigma) P_\theta^{\mathsf{FD}}(\sigma, \tau) = \mu(\tau) P_\theta^{\mathsf{FD}}(\tau, \sigma).$$

Suppose that the current configuration is $\sigma \in \{-1, +1\}^V$. By the transition rule of the field dynamics, a random $S \subseteq V$ is constructed so that each $v \in V$ with $\sigma_v = +1$ is added to S with probability θ and each $v \in V$ with $\sigma_v = -1$ is added to S with probability 1. Let $R = V \setminus S$ denote the complement of the resampled set. It must hold that $R \subseteq \sigma^{-1}(+1)$ since every variable $v \in V$ with $\sigma_v = -1$ is selected into S. To successfully transform from σ to τ, in addition, it should be satisfied that $R \subseteq \tau^{-1}(+1)$ and the resampling step generates the configuration τ_S on S. Therefore, for any $\sigma, \tau \in \Omega$,

$$(26) \qquad \mu(\sigma) P_\theta^{\mathsf{FD}}(\sigma, \tau) = \mu(\sigma) \sum_{R \subseteq \sigma^{-1}(+1) \cap \tau^{-1}(+1)} (1 - \theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \pi_S^{\mathbf{1}_R}(\tau_S)$$

$$= \mu(\sigma) \cdot \theta^{\|\sigma\|_+} \sum_{R \subseteq \sigma^{-1}(+1) \cap \tau^{-1}(+1)} \left(\frac{1 - \theta}{\theta} \right)^{|R|} \pi^{\mathbf{1}_R}(\tau) \qquad \text{(since } \tau_R = \mathbf{1}_R)$$

$$(27) \qquad = \mu(\sigma) \cdot \theta^{\|\sigma\|_+} \sum_{R \subseteq \sigma^{-1}(+1) \cap \tau^{-1}(+1)} \left(\frac{1 - \theta}{\theta} \right)^{|R|} \frac{\pi(\tau)}{\pi_R(\mathbf{1}_R)}. \qquad \text{(since } \tau_R = \mathbf{1}_R)$$

Note $\sigma \in \Omega$ is feasible with respect to both μ and π. And $\pi_R(\mathbf{1}_R) > 0$ since $\sigma_R = \mathbf{1}_R$. Thus, the conditional probability $\pi^{\mathbf{1}_R}$ in (26) and the ratio in (27) are well-defined. Recall that $\pi = \mu^{(\theta)}$. By Definition 1.7,

$$\forall \tau \in \{-1, +1\}^V, \quad \pi(\tau) = \frac{\mu(\tau) \cdot \theta^{\|\tau\|_+}}{Z(\mu, \theta)},$$

where $Z(\mu, \theta) \triangleq \sum_{\tau \in \{-1, +1\}^V} \mu(\tau) \theta^{\|\tau\|_+}$ denotes the normalizing factor that depends only on μ and θ. We have

$$\mu(\sigma) P_\theta^{\mathsf{FD}}(\sigma, \tau) = \left(\mu(\sigma) \mu(\tau) \cdot \theta^{\|\sigma\|_+ + \|\tau\|_+} \right) \sum_{R \subseteq \sigma^{-1}(+1) \cap \tau^{-1}(+1)} \left(\frac{1 - \theta}{\theta} \right)^{|R|} \frac{1}{Z(\mu, \theta) \cdot \pi_R(\mathbf{1}_R)},$$

which is symmetric in σ and τ. Therefore, the detailed balanced equation (25) holds. This concludes the proof of Proposition 2.2.

5.2. Field dynamics as the k-transformed block dynamics (proof of Lemma 2.7). Let μ be a distribution over $\{-1, +1\}^V$ and $k \in \mathbb{N}^+$. Let $\mu_k = \mathrm{Trans}(\mu, k)$ denote the k-transformation of μ defined in Definition 2.6.

We use V_k to denote the ground set for μ_k, that is,

$$(28) \qquad V_k \triangleq V \times [k].$$

For each $v \in V$ and $i \in [k]$, for convenience, we denote

$$v_i \triangleq (v, i) \in V_k.$$

Finally, for each $v \in V$, we denote

$$(29) \qquad C_v \triangleq \{v_i \mid i \in [k]\}.$$

Recall that $P_{k,\ell}$ denotes the uniform ℓ-block dynamics for $\mu_k = \mathrm{Trans}(\mu, k)$. To relate the field dynamics P_θ^{FD} with $P_{k,\ell}$, we consider the following natural projection operation which maps a configuration in the k-transformed distribution μ_k back into $\{-1, +1\}^V$.

Definition 5.1 (projection of configuration). Let μ be a distribution over $\{-1, +1\}^V$ and $k \geq 1$ an integer. Let $\mu_k = \mathrm{Trans}(\mu, k)$ denote the k-transformation of μ. For any $\sigma \in \Omega(\mu_k)$, the projection $\sigma^\star$ of σ is a configuration in $\{-1, +1\}^V$ such that for any $v \in V$

$$(30) \qquad \sigma_v^\star \triangleq \begin{cases} +1 & \text{if } \exists 1 \leq i \leq k \text{ s.t. } \sigma_{v_i} = +1 \\ -1 & \text{otherwise.} \end{cases}$$

The projection operation defined as above naturally transforms the uniform ℓ-block dynamics for the k-transformed distribution μ_k to a new stochastic process, called *projected block dynamics*, defined on the original space $\{-1, +1\}^V$.

Definition 5.2 (projected block dynamics). Let μ be a joint distribution over $\{-1, +1\}^V$, where $n = |V|$. Let $k \geq 1$ be an integer and $\mu_k = \mathrm{Trans}(\mu, k)$. Let $1 \leq \ell \leq kn$ be an integer and $(X_t)_{t \geq 0}$ the uniform ℓ-block dynamics for μ_k. The (k, ℓ)-*projected-block dynamics* $P_{k,\ell}^{\mathrm{proj}}$ is a stochastic process defined as $(X_t^\star)_{t \geq 0}$.

Not surprisingly, the stochastic process $P_{k,\ell}^{\mathrm{proj}}$ is a well-defined reversible Markov chain on space $\Omega(\mu)$ with stationary distribution μ.

Proposition 5.3. *Let μ be a distribution over $\{-1, +1\}^V$, where $n = |V|$. For all integers $k \geq 1$ and $1 \leq \ell \leq kn$, the (k, ℓ)-projected-block dynamics $P_{k,\ell}^{\mathrm{proj}}$ is a Markov chain on $\Omega(\mu)$ that is reversible with respect to μ.*

We use $\lambda_{\mathrm{gap}}(P_{k,\ell}^{\mathrm{proj}})$ to denote the spectral gap of the reversible chain $P_{k,\ell}^{\mathrm{proj}}$. We first observe that the projection in (30) applied on the uniform block dynamics $P_{k,\ell}$ does not decrease the spectral gap.

Lemma 5.4. *For all integers $k \geq 1$ and $1 \leq \ell \leq kn$,*

$$\lambda_{\mathrm{gap}}\left(P_{k,\ell}^{\mathrm{proj}}\right) \geq \lambda_{\mathrm{gap}}\left(P_{k,\ell}\right).$$

Note that the projected bock dynamics $P_{k,\ell}^{\mathrm{proj}}$ has the same state space $\Omega(\mu)$ as the field dynamics P_θ^{FD}. Therefore, we can compare their transition matrices entry-wisely. The following is a key lemma which states that the projected bock dynamics $P_{k,\ell}^{\mathrm{proj}}$ with $\ell = \lceil \theta kn \rceil$ gives an entry-wise approximation of the field dynamics P_θ^{FD} as k grows to infinity.

Lemma 5.5. *The following holds for all $0 < \theta < 1$. For any $\varepsilon > 0$, there is a finite $K = K(\mu, \theta, \varepsilon) \geq 1$ such that for all $k \geq K$, and all $X, Y \in \Omega(\mu)$,*

$$\left| P_\theta^{\mathrm{FD}}(X, Y) - P_{k, \lceil \theta kn \rceil}^{\mathrm{proj}}(X, Y) \right| \leq \varepsilon.$$

We are now ready to prove Lemma 2.7 by assuming Proposition 5.3, Lemma 5.4 and Lemma 5.5, whose proofs are postponed to Section 6.

Proof of Lemma 2.7. By Proposition 2.2, the field dynamics P_θ^{FD} is reversible with respect to distribution μ. Fix an integer $k \geq 1$. Let $\ell = \ell(k) \triangleq \lceil \theta k n \rceil$. By Proposition 5.3, the projected block dynamics $P_{k,\ell}^{\text{proj}}$ is also reversible with respect to distribution μ. Let $\Omega = \Omega(\mu)$ denote the support of μ. By the Courant-Fischer theorem [? , Lemma 13.7], we have

$$
\begin{aligned}
\lambda_{\text{gap}} \left(P_\theta^{\text{FD}} \right) &= \inf_{\substack{f \in \mathbb{R}^\Omega \\ \|f\|_{2,\mu}=1, f \perp_\mu \mathbf{1}}} \left\langle f, \left(I - P_\theta^{\text{FD}} \right) f \right\rangle_\mu \\
&= \inf_{\substack{f \in \mathbb{R}^\Omega \\ \|f\|_{2,\mu}=1, f \perp_\mu \mathbf{1}}} \left(\left\langle f, \left(I - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu - \left\langle f, \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu \right).
\end{aligned}
$$

Since μ, θ and k are all fixed, and f is a bounded function due to $\|f\|_{2,\mu} = 1$, both $\left\langle f, \left(I - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu$ and $\left\langle f, \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu$ are bounded. Therefore, the above quantity can be bounded from below as:

$$
\begin{aligned}
\lambda_{\text{gap}} \left(P_\theta^{\text{FD}} \right) &= \inf_{\substack{f \in \mathbb{R}^\Omega \\ \|f\|_{2,\mu}=1, f \perp_\mu \mathbf{1}}} \left(\left\langle f, \left(I - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu - \left\langle f, \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu \right) \\
&\geq \inf_{\substack{f \in \mathbb{R}^\Omega \\ \|f\|_{2,\mu}=1, f \perp_\mu \mathbf{1}}} \left\langle f, \left(I - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu - \sup_{\substack{f \in \mathbb{R}^\Omega \\ \|f\|_{2,\mu}=1}} \left\langle f, \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right) f \right\rangle_\mu \\
&= \lambda_{\text{gap}} \left(P_{k,\ell}^{\text{proj}} \right) - \lambda_{\max} \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right).
\end{aligned}
$$

The last equation holds because $P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}}$ is a self-adjoint operator for inner-product $\langle \cdot, \cdot \rangle_\mu$.

By Lemma 5.5, we know that for any $\varepsilon > 0$, there exists $K \geq 1$ such that for all $k \geq K$,

$$
\forall X, Y \in \Omega(\mu), \quad \left| P_\theta^{\text{FD}}(X, Y) - P_{k,\ell}^{\text{proj}}(X, Y) \right| \leq \frac{\varepsilon}{|\Omega(\mu)|},
$$

where $\ell = \ell(k) \triangleq \lceil \theta k n \rceil$. Therefore,

$$
\lambda_{\max} \left(P_\theta^{\text{FD}} - P_{k,\ell}^{\text{proj}} \right) \leq \max_{X \in \Omega(\mu)} \sum_{Y \in \Omega(\mu)} \left| P_\theta^{\text{FD}}(X, Y) - P_{k,\ell}^{\text{proj}}(X, Y) \right| \leq \varepsilon.
$$

Hence, for all $k \geq K$ and $\ell = \lceil \theta k n \rceil$,

$$
\lambda_{\text{gap}} \left(P_\theta^{\text{FD}} \right) \geq \lambda_{\text{gap}} \left(P_{k,\ell}^{\text{proj}} \right) - \varepsilon.
$$

Applying Lemma 5.4, we have that for any $\varepsilon > 0$, there exists $K \geq 1$ such that for all $k \geq K$ and $\ell = \lceil \theta k n \rceil$,

$$
\lambda_{\text{gap}} \left(P_\theta^{\text{FD}} \right) \geq \lambda_{\text{gap}} \left(P_{k,\ell} \right) - \varepsilon.
$$

Since $0 \leq \lambda_{\text{gap}} \left(P_{k,\ell} \right) \leq 2$ for all $k \geq 1$, $\limsup_{k \to \infty} \lambda_{\text{gap}} \left(P_{k,\ell} \right)$ exists. We have

$$
\lambda_{\text{gap}} \left(P_\theta^{\text{FD}} \right) \geq \limsup_{k \to \infty} \lambda_{\text{gap}} \left(P_{k,\ell} \right). \qquad \square
$$

6. Approximation of Field Dynamics

This section is dedicated to the analyses of the projected block dynamics $P^{\mathrm{proj}}_{k,\ell}$, which is an approximation of the field dynamics. In Section 6.1, we prove Proposition 5.3 for the well-defined-ness and reversibility of this projected chain; then in Section 6.2, we prove Lemma 5.4 which relates its spectral gap to the block dynamics $P_{k,\ell}$ for the k-transformed distribution μ_k; and finally in Section 6.3, we prove Lemma 5.5 for its entry-wise approximation of the field dynamics P^{FD}_θ. Altogether, they imply Lemma 2.7.

6.1. Well-definedness of projected dynamics (proof of Proposition 5.3).

We show that the projected block dynamics in Definition 5.2 is precisely the following Markov chain $\mathcal{M}$. The chain $\mathcal{M}$ starts from an arbitrary $X \in \Omega(\mu)$. In each transition step, the current configuration $X \in \Omega(\mu)$ is updated as:

- Sample $\boldsymbol{a} = (a_v)_{v \in V}$ according to the multivariate hypergeometric distribution $\Pi_{V,k,\ell}$, and let $\boldsymbol{b} = \frac{\boldsymbol{a}}{k}$.
- Construct a random $S \subseteq V$ by independently selecting each $v \in V$ into S with probability

$$q_v = \begin{cases} 1 & \text{if } X_v = -1, \\ b_v & \text{if } X_v = +1. \end{cases}$$

- Replace X_S by a random partial configuration sampled according to $\mu^{(\boldsymbol{b},S),X_{V\setminus S}}_S$, where $\mu^{(\boldsymbol{b},S),X_{V\setminus S}}_S$ denotes the marginal distribution on S induced from $\mu^{(\boldsymbol{\phi})}$ conditional on $X_{V\setminus S}$, where $\boldsymbol{\phi} \in \mathbb{R}^V_{\geq 0}$ is defined as

$$(31) \qquad \forall v \in V, \quad \phi_v = \begin{cases} b_v & \text{if } v \in S, \\ 1 & \text{if } v \notin S. \end{cases}$$

Note that there may exist a subset $H \subseteq V$ such that $\phi_H = \boldsymbol{0}_H$, where $\boldsymbol{0}_H$ denotes the all-zero vector on H. And by definition of $\mu^{(\boldsymbol{\phi})}$, this is equivalent to conditioning on the configuration on H being fixed as $(-1)_H$, which is the all-(-1) configuration on H. For all such $v \in H$, it must hold that $v \in S$ and $b_v = 0$, which means $X_v = -1$. Since $X \in \Omega(\mu)$, we have $\mu_H((-1)_H) > 0$. Hence, $\mu^{(\boldsymbol{\phi})}$ is well-defined. It is straightforward to verify X is feasible with respect to $\mu^{(\boldsymbol{\phi})}$, thus $\mu^{(\boldsymbol{b},S),X_{V\setminus S}}_S$ is also well-defined.

Lemma 6.1. *The Markov chain $\mathcal{M}$ is precisely the projected block dynamics $P^{\mathrm{proj}}_{k,\ell}$. Formally, for all configurations $X, Y \in \Omega(\mu)$, we have the following identity:*

$$(32) \qquad \forall \sigma \in \Omega(\mu_k) \text{ that } \sigma^\star = X, \quad \mathcal{M}(X,Y) = \sum_{\tau \in \Omega(\mu_k):\tau^\star = Y} P_{k,\ell}(\sigma, \tau).$$

The identity (32) automatically confirms the Markovian property of $P^{\mathrm{proj}}_{k,\ell}$, because it confirms that in the pre-projection chain $P_{k,\ell}$, the transition probability $\sum_{\tau \in \Omega(\mu_k):\tau^\star = Y} P_{k,\ell}(\sigma, \tau)$ from any $\sigma \in \Omega(\mu_k)$ to the class of configurations $\tau \in \Omega(\mu_k)$ projected to the same $\tau^\star = Y$, is constant for all $\sigma \in \Omega(\mu_k)$ that $\sigma^\star = X$.

Assuming that Lemma 6.1 holds, we can prove Proposition 5.3.

Proof of Proposition 5.3 assuming Lemma 6.1. It suffices to show that the Markov chain $\mathcal{M}$ is reversible with respect to μ. Fix two feasible configurations $X, Y \in \Omega(\mu)$. Recall that $X^{-1}(+1) \triangleq \{v \in V \mid X_v = +1\}$ denotes the pre-image of $+1$ under X. In each transition step, the chain $\mathcal{M}$ first generates a vector $\boldsymbol{b}$ with probability $\Pi_{V,k,\ell}(k\boldsymbol{b})$, and then samples a random subset $S \subseteq V$. Consider $R = V \setminus S$. To transform from

X to Y, it must hold that $R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)$. We denote $\overline{R} \triangleq V \setminus R = S$. The following equation holds:

$$\mathcal{M}(X, Y) = \sum_{\boldsymbol{b}:k\boldsymbol{b}\in\Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(k\boldsymbol{b}) \sum_{R\subseteq X^{-1}(+1)\cap Y^{-1}(+1)} \left(\prod_{v\in X^{-1}(+1)\setminus R} b_v \prod_{v\in R}(1-b_v) \right) \mu_{\overline{R}}^{(\boldsymbol{b},\overline{R}),\mathbf{1}_R}(Y_{\overline{R}})$$

$$\tag{33} = \sum_{\boldsymbol{b}:k\boldsymbol{b}\in\Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(k\boldsymbol{b}) \sum_{R\subseteq X^{-1}(+1)\cap Y^{-1}(+1)} \left(\prod_{v\in X^{-1}(+1)\setminus R} b_v \prod_{v\in R}(1-b_v) \right) \mu^{(\boldsymbol{b},\overline{R}),\mathbf{1}_R}(Y),$$

where the last equation holds because $Y_R = \mathbf{1}_R$. Also due to $Y_R = \mathbf{1}_R$, we have

$$\mu^{(\boldsymbol{b},\overline{R}),\mathbf{1}_R}(Y) = \frac{\mu^{(\boldsymbol{b},\overline{R})}(Y)}{\mu_R^{(\boldsymbol{b},\overline{R})}(\mathbf{1}_R)}.$$

By definition of $\mu^{(\boldsymbol{b},\overline{R})} = \mu^{(\boldsymbol{\phi})}$ where $\boldsymbol{\phi} \in \mathbb{R}_{\geq 0}^V$ is as defined in (31), we have

$$\mu^{(\boldsymbol{b},\overline{R})}(Y) = \frac{1}{Z}\mu(Y) \prod_{v\in Y^{-1}(+1)\setminus R} b_v,$$

where $Z = Z(\mu, \boldsymbol{b}, R) \triangleq \sum_{\sigma\in\Omega(\mu)} \mu(\sigma) \prod_{v\in\sigma^{-1}(+1)\setminus R} b_v$. Hence, $\mu(X)\mathcal{M}(X, Y)$ can be expressed as

$$\frac{\mu(X)\mu(Y)}{Z} \sum_{\boldsymbol{b}:k\boldsymbol{b}\in\Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(k\boldsymbol{b}) \sum_{R\subseteq X^{-1}(+1)\cap Y^{-1}(+1)} \left(\frac{1}{\mu_R^{(\boldsymbol{b},\overline{R})}(\mathbf{1}_R)} \prod_{v\in X^{-1}(+1)\setminus R} b_v \prod_{v\in Y^{-1}(+1)\setminus R} b_v \prod_{v\in R}(1-b_v) \right),$$

which is symmetric in X and Y. Therefore, the detailed balance equation is satisfied:

$$\mu(X)\mathcal{M}(X, Y) = \mu(Y)\mathcal{M}(Y, X).$$

The chain $\mathcal{M}$ is reversible with respect to μ. $\qquad\square$

Recall that we use $V_k \triangleq V \times [k]$ to denote the ground set of μ_k. For each $v \in V$ and $i \in [k]$, we denote $v_i \triangleq (v, i) \in V_k$ and $C_v \triangleq \{v_i \mid i \in [k]\}$.

Let $\Lambda \subseteq V_k$ and $\rho \in \Omega(\mu_{k,\Lambda})$. Define

$$F(\rho) \triangleq \{v \in V \mid \exists i \in [k] \text{ s.t. } v_i \in \Lambda \text{ and } \rho_{v_i} = +1\}.$$

Define the local fields $\boldsymbol{\phi}_\rho$ specified by $\rho \in \Omega(\mu_{k,\Lambda})$ as

$$\forall v \in V, \quad \phi_\rho(v) \triangleq \begin{cases} \frac{|C_v\setminus\Lambda|}{k}, & \text{if } v \notin F(\rho) \\ 1 & \text{if } v \in F(\rho). \end{cases}$$

We need the following lemma to prove Lemma 6.1.

Lemma 6.2. *For all $\Lambda \subseteq V_k$, $\rho \in \Omega(\mu_{k,\Lambda})$, and $\xi \in \Omega(\mu)$, the distribution $\mu^{(\boldsymbol{\phi}_\rho),\mathbf{1}_{F(\rho)}}$ is well-defined and*

$$\mu^{(\boldsymbol{\phi}_\rho),\mathbf{1}_{F(\rho)}}(\xi) = \sum_{\substack{\tau\in\Omega(\mu_k):\\ \tau^\star=\xi}} \mu_k^\rho(\tau),$$

where $\mu^{(\boldsymbol{\phi}_\rho),\mathbf{1}_{F(\rho)}}$ is induced from $\mu^{(\boldsymbol{\phi}_\rho)}$ conditional on $\mathbf{1}_{F(\rho)}$, the all-(+1) configuration on $F(\rho)$.

We first use Lemma 6.2 to prove Lemma 6.1, and then prove Lemma 6.2.

Proof of Lemma 6.1. It suffices to verify the identity in (32).

Consider the transition step $\sigma \to \tau$ in the block dynamics. The block dynamics pick a subset $S \in \binom{V_k}{\ell}$ uniformly at random. Observe that the following two processes are equivalent:

- Pick a subset $S \in \binom{V_k}{\ell}$ uniformly at random.

- Sample $a = (a_v)_{v \in V}$ according to the multivariate hypergeometric distribution $\Pi_{V,k,\ell}$; for each $v \in V$, pick a subset $S_v \in \binom{C_v}{a_v}$ uniformly at random; and finally let $S = \cup_{v \in V} S_v$.

Imagine that we use the second process to pick the subset $S = \cup_{v \in V} S_v$. We have

$$P_{k,\ell}(\sigma, \tau) = \sum_{a \in \Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(a) \sum_{\substack{(S_v)_{v \in V}: \\ S_v \subseteq C_v \text{ and } |S_v| = a_v}} \left(\prod_{v \in V} \binom{k}{a_v}^{-1} \right) \mu_k^{\sigma_{V_k \setminus S}}(\tau), \text{ where } S = \cup_{v \in V} S_v$$

$$(34) \qquad = \mathbf{E}_{a \sim \Pi_{V,k,\ell}} \left[\mathbf{E}_{\substack{(S_v)_{v \in V}: \\ S_v \sim_U \binom{C_v}{a_v}}} \left[\mu_k^{\sigma_{V_k \setminus S}}(\tau) \right] \right], \text{ where } S = \cup_{v \in V} S_v,$$

where the expectation is taken over the random choices of $(S_v)_{v \in V}$, where each $S_v \sim_U \binom{C_v}{a_v}$ is sampled from $\binom{C_v}{a_v}$ uniformly and independently. Recall that the set $F(\sigma) \subseteq V$ is defined as

$$(35) \qquad F(\sigma) \triangleq \{v \in V \mid \exists i^* \in [k] \text{ s.t. } v_{i^*} \in V_k \text{ and } \sigma_{v_{i^*}} = +1\}.$$

Note that for all $v \in V$, the index $i^* \in [k]$ is unique. Since $\sigma^\star = X$, it is straightforward to verify that

$$X^{-1}(+1) = F(\sigma).$$

Given a sequence of sets $(S_v)_{v \in V}$, where $S_v \subseteq C_v$, we can define a subset of $X^{-1}(+1)$ as follows

$$R_{(S_v)_{v \in V}} \triangleq \{v \in X^{-1}(+1) \mid v_{i^*} \notin S_v\}.$$

where $i^* = i^*(v) \in [k]$ is the unique index for v in (35). In other words, R is the subset of $X^{-1}(+1)$ satisfying that the unique variable v_{i^*} with $\sigma_{v_{i^*}} = +1$ is not picked by set S. Given $R \subseteq X^{-1}(+1)$ and a sequence of sets $(S_v)_{v \in V}$, where $S_v \subseteq C_v$, we say $(S_v)_{v \in V}$ is *consistent* with R if $R = R_{(S_v)_{v \in V}}$. By (34),

$$P_{k,\ell}(\sigma, \tau) = \mathbf{E}_{a \sim \Pi_{V,k,\ell}} \left[\sum_{R \subseteq X^{-1}(+1)} \mathbf{E}_{\substack{(S_v)_{v \in V}: \\ S_v \sim_U \binom{C_v}{a_v}}} \left[\mu_k^{\sigma_{V_k \setminus S}}(\tau) \mathbf{1}[R = R_{(S_v)_{v \in V}}] \right] \right],$$

where $S = \cup_{v \in V} S_v$. Fix any $Y \in \Omega(\mu)$. To prove (32), we need to calculate

$$\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = Y}} P_{k,\ell}(\sigma, \tau) = \mathbf{E}_{a \in \Pi_{V,k,\ell}} \left[\sum_{R \subseteq X^{-1}(+1)} \mathbf{E}_{\substack{(S_v)_{v \in V}: \\ S_v \sim_U \binom{C_v}{a_v}}} \left[\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = Y}} \mu_k^{\sigma_{V_k \setminus S}}(\tau) \mathbf{1}[R = R_{(S_v)_{v \in V}}] \right] \right].$$

To simplify the above equation, we need the following result. Consider the equation (34). We fix $a \in \Omega(\Pi_{V,k,\ell})$ and $R \subseteq X^{-1}(+1)$. We claim the following result holds: For any sequence $(S_v)_{v \in V}$, where $S_v \subseteq C_v$, if the following two conditions hold together:

- for all $v \in V$, $|S_v| = a_v$,
- $(S_v)_{v \in V}$ is consistent with R,

then for $S = \cup_{v \in V} S_v$, and for all $Y \in \Omega(\mu)$, it holds that

$$(36) \qquad \sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = Y}} \mu_k^{\sigma_{V_k \setminus S}}(\tau) = \mu^{(\varphi_{a,R}), \mathbf{1}_R}(Y),$$

where $\varphi_{a,R}$ is defined as

$$\forall v \in V, \quad \varphi_{a,R}(v) \triangleq \begin{cases} \frac{a_v}{k} & \text{if } v \notin R, \\ 1 & \text{if } v \in R. \end{cases}$$

Equation (36) follows from Lemma 6.2. This is because $F(\sigma_{V_k \setminus S}) = R$ and $C_v \setminus (V_k \setminus S) = C_v \cap S$ and $|C_v \cap S| = |S_v| = a_v$.

Hence, we have the following equation

$$\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = Y}} P_{k,\ell}(\sigma, \tau) = \mathbf{E}_{a \in \Pi_{V,k,\ell}} \left[\sum_{R \subseteq X^{-1}(+1)} \mu^{(\varphi_{a,R}),\mathbf{1}_R}(Y) \mathbf{E}_{\substack{(S_v)_{v \in V}: \\ S_v \sim U\binom{C_v}{a_v}}} \left[\mathbf{1}[R = R_{(S_v)_{v \in V}}] \right] \right]$$

$$= \mathbf{E}_{a \in \Pi_{V,k,\ell}} \left[\sum_{R \subseteq X^{-1}(+1)} \mu^{(\varphi_{a,R}),\mathbf{1}_R}(Y) \mathbf{Pr}_{\substack{(S_v)_{v \in V}: \\ S_v \sim U\binom{C_v}{a_v}}} \left[R_{(S_v)_{v \in V}} = R \right] \right]$$

$$= \sum_{a \in \Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(a) \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} \left(\frac{a_v}{k} \right) \prod_{v \in R} \left(1 - \frac{a_v}{k} \right) \right) \mu^{(\varphi_{a,R}),\mathbf{1}_R}(Y),$$

where the last equation holds because for each $v \in V$ and $i \in [k]$, probability of event $v_i \in S_v$ is $\frac{a_i}{k}$ and $\mu^{(\varphi_{a,R}),\mathbf{1}_R}(Y) = 0$ if $Y_R \neq \mathbf{1}_R$. Recall that $b = a/k$. By (33), we have

$$\mathcal{M}(X, Y) = \sum_{a \in \Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(a) \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} \left(\frac{a_v}{k} \right) \prod_{v \in R} \left(1 - \frac{a_v}{k} \right) \right) \mu^{(a/k,\overline{R}),\mathbf{1}_R}(Y),$$

where $\overline{R} = V \setminus R$. By the definition in (31), $\mu^{(a/k,\overline{R})}$ is obtained from μ by imposing the local fields ϕ, where

$$\forall v \in V, \quad \phi_v = \begin{cases} b_v = \frac{a_v}{k} & \text{if } v \notin R \\ 1 & \text{if } v \in R. \end{cases}$$

Hence, $\mu^{(a/k,\overline{R})}$ and $\mu^{(\varphi_{a,R})}$ are the same distribution, which implies

$$\mathcal{M}(X, Y) = \sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = Y}} P_{k,\ell}(\sigma, \tau).$$

This proves (32). □

Proof of Lemma 6.2. For any $v \in V$ such that $\phi_\rho(v) = 0$, it must hold that $v \notin F(\rho)$ and $|C_v \setminus \Lambda| = 0$, which implies $C_v \subseteq \Lambda$ and $\rho_{v_i} = -1$ for all $i \in [k]$. Therefore, for any $\sigma \in \Omega(\mu_k^\rho)$, $\sigma^\star$ (defined in (30)) is a feasible configuration of $\mu^{(\phi_\rho),\mathbf{1}_{F(\rho)}}$. This guarantees that $\mu^{(\phi_\rho),\mathbf{1}_{F(\rho)}}$ is a well-defined distribution.

Fix $\xi \in \Omega(\mu)$. By definition of distribution μ_k, we have

$$\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi}} \mu_k^\rho(\tau) = \sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi}} \frac{\mathbf{Pr}_{Y \sim \mu_k}[Y_\Lambda = \rho \text{ and } Y = \tau]}{\mathbf{Pr}_{Y \sim \mu_k}[Y_\Lambda = \rho]} = \sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi, \tau_\Lambda = \rho}} \frac{\mathbf{Pr}_{Y \sim \mu_k}[Y = \tau]}{\mathbf{Pr}_{Y \sim \mu_k}[Y_\Lambda = \rho]}.$$

Note that if $\tau^\star = \xi$, then it holds that $\|\tau\|_{+1} = \|\xi\|_{+1}$. It holds that

$$\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi, \tau_\Lambda = \rho}} \mathbf{Pr}_{Y \sim \mu_k}[Y = \tau] = \mu(\xi) \left(\frac{1}{k} \right)^{\|\xi\|_{+1}} \sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi}} \mathbf{1}[\tau_\Lambda = \rho]$$

$$= \mu(\xi) \left(\frac{1}{k} \right)^{\|\xi\|_{+1}} \mathbf{1}[\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}] \prod_{v \in \xi^{-1}(+1) \setminus F(\rho)} |C_v \setminus \Lambda|.$$

The last equation holds because:

- if $\tau^\star = \xi$ and $\tau_\Lambda = \rho$, then it must hold that $\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}$;

24

- for all $v \in \xi^{-1}(+1)$, we need to choose one index $i \in [k]$ and set $\tau_{v_i} = +1$, if $v \in F(\rho)$ such index is fixed by ρ, if $v \in \xi^{-1}(+1) \setminus F(\rho)$, there are $|C_v \setminus \Lambda|$ ways to choose such index.

Reorganizing above equation gives

$$\sum_{\substack{\tau \in \Omega(\mu_k): \\ \tau^\star = \xi, \tau_\Lambda = \rho}} \mathbf{Pr}_{Y \sim \mu_k}[Y = \tau] = \mu(\xi) \mathbf{1}[\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}] \left(\frac{1}{k}\right)^{|F(\rho)|} \prod_{v \in \xi^{-1}(+1) \setminus F(\rho)} \frac{|C_v \setminus \Lambda|}{k}.$$

Next, we have

$$\mathbf{Pr}_{Y \sim \mu_k}[Y_\Lambda = \rho] = \sum_{\eta \in \Omega(\mu)} \mu(\eta) \mathbf{1}[\eta_{F(\rho)} = \mathbf{1}_{F(\rho)}] \left(\frac{1}{k}\right)^{|F(\rho)|} \prod_{v \in \eta^{-1}(+1) \setminus F(\rho)} \frac{|C_v \setminus \Lambda|}{k}.$$

By definition, $Y \sim \mu_k$ is obtained by first sampling $X \sim \mu$, then transforming X to Y. The above equation enumerates all values η for X. The event $Y_\Lambda = \rho$ occurs if and only if:

- $\eta_v = +1$ for all $v \in F(\rho)$;
- for all $v \in F(\rho)$, there is a unique index $i \in [k]$ fixed by ρ such that $v_i = +1$ (this event occurs with probability $\left(\frac{1}{k}\right)^{|F(\rho)|}$);
- for all other $v \in \eta^{-1}(+1) \setminus F(\rho)$, we choose an index $i \in [k]$ to set $v_i = +1$ and $v_i \notin \Lambda$ (this event occurs with probability $\prod_{v \in \eta^{-1}(+1) \setminus F(\rho)} \frac{|C_v \setminus \Lambda|}{k}$).

Combining above two equations together, we have

$$\sum_{\tau \in \Omega(\mu_k): \tau^\star = \xi} \mu_k^\rho(\tau) = \sum_{\tau \in \Omega(\mu_k): \tau^\star = \xi, \tau_\Lambda = \rho} \frac{\mathbf{Pr}_{Y \sim \mu_k}[Y = \tau]}{\mathbf{Pr}_{Y \sim \mu_k}[Y_\Lambda = \rho]}$$

$$= \frac{\mu(\xi) \mathbf{1}[\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}] \prod_{v \in \xi^{-1}(+1) \setminus F(\rho)} \frac{|C_v \setminus \Lambda|}{k}}{\sum_{\eta \in \Omega(\mu)} \mu(\eta) \mathbf{1}[\eta_{F(\rho)} = \mathbf{1}_{F(\rho)}] \prod_{v \in \eta^{-1}(+1) \setminus F(\rho)} \frac{|C_v \setminus \Lambda|}{k}}$$

$$= \frac{\mathbf{1}[\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}] \mu^{(\phi_\rho)}(\xi)}{\sum_{\eta \in \Omega(\mu)} \mathbf{1}[\eta_{F(\rho)} = \mathbf{1}_{F(\rho)}] \mu^{(\phi_\rho)}(\eta)}$$

$$= \frac{\mathbf{1}[\xi_{F(\rho)} = \mathbf{1}_{F(\rho)}] \mu^{(\phi_\rho)}(\xi)}{\sum_{\eta \in \Omega(\mu^{(\phi_\rho)})} \mathbf{1}[\eta_{F(\rho)} = \mathbf{1}_{F(\rho)}] \mu^{(\phi_\rho)}(\eta)}$$

$$= \mu^{(\phi_\rho), \mathbf{1}_{F(\rho)}}(\xi),$$

where the second to the last equation holds because $\Omega(\mu^{(\phi_\rho)}) \subseteq \Omega(\mu)$. This proves the lemma. □

6.2. **Comparing projected block dynamics with block dynamics (proof of Lemma 5.4).** By Proposition 5.3, $P_{k,\ell}^{\mathrm{proj}}$ is a reversible Markov chain. We establish the following lemma which implies Lemma 5.4.

Lemma 6.3. *For any function $f : \Omega(\mu) \to \mathbb{R}$ satisfying $\mathbf{Var}_\mu[f] \neq 0$, there exists a function $f' : \Omega(\mu_k) \to \mathbb{R}$ such that $\mathbf{Var}_{\mu_k}[f'] \neq 0$ and*

$$\frac{\mathcal{E}_{P_{k,\ell}^{\mathrm{proj}}}(f, f)}{\mathbf{Var}_\mu(f)} = \frac{\mathcal{E}_{P_{k,\ell}}(f', f')}{\mathbf{Var}_{\mu_k}(f')}.$$

With this lemma, Lemma 5.4 follows immediately from the characterization of spectral gap in (14):

$$\lambda_{\mathrm{gap}}(P) = \inf_{\substack{f \in \mathbb{R}^{\Omega(\mu)} \\ \mathbf{Var}_\mu[f] \neq 0}} \frac{\mathcal{E}_P(f, f)}{\mathbf{Var}_\mu[f]}.$$

Now, we prove Lemma 6.3.

25

Proof of Lemma 6.3. Given a function $f : \Omega(\mu) \to \mathbb{R}$ satisfying $\mathbf{Var}_\mu[f] \neq 0$, the function $f' : \Omega(\mu_k) \to \mathbb{R}$ is constructed as follows:

$$(37) \qquad \forall \xi \in \Omega(\mu_k), \quad f'(\xi) = f(\xi^\star),$$

where $\xi^\star$ is as defined in (30).

Next, we prove the following identities that implies the lemma:

$$\mathcal{E}_{P_{k,\ell}}(f', f') = \mathcal{E}_{P_{k,\ell}^{\mathrm{proj}}}(f, f),$$

$$\mathbf{Var}_{\mu_k}(f') = \mathbf{Var}_\mu(f).$$

For any $X, Y \in \Omega(\mu)$, by the definition of μ_k, it holds that

$$(38) \qquad \forall \sigma \in \Omega(\mu_k) \text{ satisfying } \sigma^\star = X, \quad \mu(\sigma) = \sum_{\substack{\xi \in \Omega(\mu_k) \\ \xi^\star = \sigma}} \mu_k(\xi).$$

By (32), it holds that

$$(39) \qquad \forall \sigma \in \Omega(\mu_k) \text{ satisfying } \sigma^\star = X, \quad P_{k,\ell}^{\mathrm{proj}}(X, Y) = \sum_{\substack{\tau \in \Omega(\mu_k) \\ \tau^\star = Y}} P_{k,\ell}(\sigma, \tau).$$

Therefore, we have

$$\mathcal{E}_{P_{k,\ell}}(f', f') = \frac{1}{2} \sum_{\sigma, \tau \in \Omega(\mu_k)} \mu_k(\sigma) P_{k,\ell}(\sigma, \tau) (f'(\sigma) - f'(\tau))^2$$

$$= \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} \sum_{\substack{\sigma, \tau \in \Omega(\mu_k) \\ \sigma^\star = X, \tau^\star = Y}} \mu_k(\sigma) P_{k,\ell}(\sigma, \tau) (f'(\sigma) - f'(\tau))^2$$

$$\text{(by definition of } f') \quad = \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} \sum_{\substack{\sigma, \tau \in \Omega(\mu_k) \\ \sigma^\star = X, \tau^\star = Y}} \mu_k(\sigma) P_{k,\ell}(\sigma, \tau) (f(X) - f(Y))^2$$

$$= \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} (f(X) - f(Y))^2 \sum_{\substack{\sigma \in \Omega(\mu_k) \\ \sigma^\star = X}} \mu_k(\sigma) \sum_{\substack{\tau \in \Omega(\mu_k) \\ \tau^\star = Y}} P_{k,\ell}(\sigma, \tau)$$

$$\text{(by (38) and (39))} \quad = \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} \mu(X) P_{k,\ell}^{\mathrm{proj}}(X, Y) (f(X) - f(Y))^2 = \mathcal{E}_{P_{k,\ell}^{\mathrm{proj}}}(f, f).$$

Similarly,

$$\mathbf{Var}_{\mu_k}[f'] = \frac{1}{2} \sum_{\sigma, \tau \in \Omega(\mu_k)} \mu_k(\sigma) \mu_k(\tau) (f'(\sigma) - f'(\tau))^2$$

$$= \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} (f(X) - f(Y))^2 \sum_{\substack{\sigma \in \Omega(\mu_k) \\ \sigma^\star = X}} \mu_k(\sigma) \sum_{\substack{\tau \in \Omega(\mu_k) \\ \tau^\star = Y}} \mu_k(\tau)$$

$$\text{(by (38))} \quad = \frac{1}{2} \sum_{X, Y \in \Omega(\mu)} \mu(X) \mu(Y) (f(X) - f(Y))^2 = \mathbf{Var}_\mu[f].$$

This proves the lemma. $\qquad \square$

6.3. Approximation of the field dynamics (proof of Lemma 5.5). Let $n = |V|$. Fix a real number $0 < \theta < 1$. Given $\varepsilon > 0$, we show that for any $X, Y \in \Omega(\mu)$, there is a $K = K(X, Y, \mu, \theta, \varepsilon)$ such that for any $k \geq K$,

$$(40) \qquad \left| P_\theta^{\mathrm{FD}}(X, Y) - P_{k, \lceil \theta k n \rceil}^{\mathrm{proj}}(X, Y) \right| \leq \varepsilon.$$

Note that $\Omega(\mu)$ is independent of k. Lemma 5.5 follows by taking $K(\mu, \theta, \varepsilon) = \max_{X, Y \in \Omega(\mu)} K(X, Y, \mu, \theta, \varepsilon)$.
   Fix $X, Y \in \Omega(\mu)$ and integer $k \geq 1$. Let $\ell \triangleq \lceil \theta n k \rceil$. By Lemma 6.1 and Equation (33),

$$(41) \quad P_{k, \ell}^{\mathrm{proj}}(X, Y) = \sum_{a : a \in \Omega(\Pi_{V, k, \ell})} \Pi_{V, k, \ell}(a) \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} \frac{a_v}{k} \prod_{v \in R} \left(1 - \frac{a_v}{k} \right) \right) \mu^{(a/k, \overline{R}), \mathbf{1}_R}(Y),$$

where $\mu^{(a/k, \overline{R})}$ is distribution $\mu^{(\phi)}$ for local fields $\phi \in \mathbb{R}_{\geq 0}^V$ defined by

$$\forall v \in V, \quad \phi_v = \begin{cases} \frac{a_v}{k} & \text{if } v \notin R \\ 1 & \text{if } v \in R. \end{cases}$$

Let $f_\mu(w) : (0, 1)^V \to \mathbb{R}$ be a function defined as follows:

$$f_\mu(w) = \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} w_v \prod_{v \in R} (1 - w_v) \right) \mu^{(w, \overline{R}), \mathbf{1}_R}(Y)$$

$$= \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} w_v \prod_{v \in R} (1 - w_v) \right) \frac{\mu(Y) \prod_{v \in Y^{-1}(+1) \setminus R} w_v}{\sum_{Z \in \Omega(\mu) : Z_R = \mathbf{1}_R} \mu(Z) \prod_{v \in Z^{-1}(+1) \setminus R} w_v}$$

Since $Y \in \Omega(\mu)$ and $Y_R = \mathbf{1}_R$, for all $w \in (0, 1)^V$, it holds that $\sum_{Z \in \Omega(\mu) : Z_R = \mathbf{1}_R} \mu(Z) \prod_{v \in Z^{-1}(+1) \setminus R} w_v > 0$. The function $f_\mu(w)$ is a rational function. Thus $f_\mu(w)$ is continuous on $(0, 1)^V$. Formally, we have:

Fact 6.4. *For any $\varepsilon > 0$ and $w \in (0, 1)^V$, there exists a constant $\delta = \delta(\mu, w) > 0$ such that for all $w' \in (0, 1)^V$ satisfying $\|w - w'\|_\infty < \delta$, it holds that*

$$\left| f_\mu(w) - f_\mu(w') \right| < \varepsilon,$$

where $\|w - w'\|_\infty \triangleq \max_{v \in V} |w(v) - w'(v)|$ is the infinity norm of $w - w'$.

   We now prove (40). Fix a parameter $\varepsilon > 0$. Let $\theta \triangleq (\theta)_{v \in V} = \theta \mathbf{1}$. By Fact 6.4, there exists a $\delta = \delta(\mu, \theta, \varepsilon) \in \left(0, \frac{\min\{\theta, 1 - \theta\}}{2} \right)$ such that for all $w \in (0, 1)^V$ satisfying $\|w - \theta\|_\infty < \delta$

$$(42) \qquad \left| f_\mu(w) - f_\mu(\theta) \right| < \frac{\varepsilon}{2}.$$

Define a subset of $\mathcal{B}_{\delta, k} \subseteq \Omega(\Pi_{V, k, \ell})$ by

$$(43) \qquad \mathcal{B}_{\delta, k} = \left\{ a \in \Omega(\Pi_{V, k, \ell}) \mid \left\| \frac{a}{k} - \theta \right\|_\infty \geq \delta \right\}.$$

Note that the function f_μ satisfies

$$\forall w \in (0, 1)^V, \quad 0 < f_\mu(w) \leq \sum_{R \subseteq X^{-1}(+1) \cap Y^{-1}(+1)} \left(\prod_{v \in X^{-1}(+1) \setminus R} w_v \prod_{v \in R} (1 - w_v) \right) \leq 1.$$

Combining (42) and (43), for any $\boldsymbol{a} \notin \mathcal{B}_{\delta,k}$, since $\delta \in \left(0, \frac{\min\{\theta, 1-\theta\}}{2}\right)$, $\boldsymbol{a}$ is a positive vector and it holds that $\left|f_\mu(\boldsymbol{a}/k) - f_\mu(\boldsymbol{\theta})\right| < \frac{\varepsilon}{2}$. By (41), on the one hand,

$$
\begin{aligned}
P_{k,\ell}^{\mathrm{proj}}(X, Y) &\geq \sum_{\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell}) \setminus \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) f_\mu\left(\frac{\boldsymbol{a}}{k}\right) \\
&\geq \left(f_\mu(\boldsymbol{\theta}) - \frac{\varepsilon}{2}\right) \sum_{\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell}) \setminus \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) \\
&= \left(f_\mu(\boldsymbol{\theta}) - \frac{\varepsilon}{2}\right) \left(1 - \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right]\right).
\end{aligned}
$$

(44)

This implies that

(45)
$$
P_{k,\ell}^{\mathrm{proj}}(X, Y) \geq f_\mu(\boldsymbol{\theta}) - \frac{\varepsilon}{2} - \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right].
$$

This is because (45) holds trivially if $f_\mu(\boldsymbol{\theta}) \leq \frac{\varepsilon}{2}$, and if otherwise $f_\mu(\boldsymbol{\theta}) > \frac{\varepsilon}{2}$, since $f_\mu(\boldsymbol{\theta}) \leq 1$, (45) is a consequence of (44).

On the other hand,

$$
\begin{aligned}
P_{k,\ell}^{\mathrm{proj}}(X, Y) &= \sum_{\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell})} \Pi_{V,k,\ell}(\boldsymbol{a}) f_\mu\left(\frac{\boldsymbol{a}}{k}\right) \\
&= \sum_{\boldsymbol{a} \in \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) f_\mu\left(\frac{\boldsymbol{a}}{k}\right) + \sum_{\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell}) \setminus \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) f_\mu\left(\frac{\boldsymbol{a}}{k}\right) \\
&\leq \sum_{\boldsymbol{a} \in \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) + \left(f_\mu(\boldsymbol{\theta}) + \frac{\varepsilon}{2}\right) \sum_{\boldsymbol{a} \in \Omega(\Pi_{V,k,\ell}) \setminus \mathcal{B}_{\delta,k}} \Pi_{V,k,\ell}(\boldsymbol{a}) \\
&= \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right] + \left(f_\mu(\boldsymbol{\theta}) + \frac{\varepsilon}{2}\right) \left(1 - \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right]\right) \\
&\leq \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right] + f_\mu(\boldsymbol{\theta}) + \frac{\varepsilon}{2},
\end{aligned}
$$

(46)

Finally, we also claim that there is a $K = K(X, Y, \mu, \theta, \varepsilon)$ such that for any $k \geq K$

(47)
$$
\mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right] \leq \frac{\varepsilon}{2}.
$$

Altogether, (45), (46) and (47) implies (41).

It remains to prove (47). Recall that $n = |V|$ and $\ell = \lceil \theta k n \rceil$. Observe that when $k \geq \frac{2}{\delta}$,

$$
\begin{aligned}
\mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\boldsymbol{a} \in \mathcal{B}_{\delta,k}\right] &= \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\exists v \in V, \left|\frac{a_v}{k} - \theta\right| \geq \delta\right] \\
&\leq \sum_{v \in V} \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\left|\frac{a_v}{k} - \theta\right| \geq \delta\right] \\
&\leq \sum_{v \in V} \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\left|\frac{a_v}{k} - \frac{\lceil \theta k n \rceil}{kn}\right| \geq \frac{\delta}{2}\right],
\end{aligned}
$$

(48)

where the last inequality holds for $k \geq \frac{2}{\delta}$.

Furthermore, by Lemma 3.9, there exists a constant $K_0 = K_0(n, \delta)$ such that for $k \geq K_0$,

(49)
$$
\forall v \in V, \quad \mathbf{Pr}_{\boldsymbol{a} \sim \Pi_{V,k,\ell}} \left[\left|\frac{a_v}{k} - \frac{\lceil \theta k n \rceil}{kn}\right| \geq \frac{\delta}{2}\right] \leq 2 \exp\left(\frac{-\delta^2 k}{2}\right) \leq \frac{\varepsilon}{2n}.
$$

Note that $\delta = \delta(\mu, \theta, \varepsilon)$ and n is determined by μ. Combining (48) and (49) proves (47).

7. Mixing of Block Dynamics

In this section, we prove Lemma 2.8 for the mixing of the uniform block dynamics for the k-transformed distribution $\mu_k = \mathsf{Trans}(\mu, k)$, assuming the complete spectral independence of μ. Together with Lemma 2.7 proved in the last two sections, this proves the mixing lemma for the field dynamics (Lemma 2.3)

With the Chen-Liu-Vigoda theorem (Theorem 2.5), to prove Lemma 2.8, we only need to verify that the k-transformed distribution μ_k is spectrally independent knowing that the original distribution μ is completely spectrally independent.

Lemma 7.1. *Let μ be a distribution over $\{-1, +1\}^V$ and $\eta > 0$. If μ is completely η-spectrally independent, then for all integers $k \geq 1$, $\mu_k = \mathsf{Trans}(\mu, k)$ is $(\eta + 2)$-spectrally independent.*

Lemma 2.8 is an easy consequence of Theorem 2.5 and Lemma 7.1.

Our remaining task is to prove Lemma 7.1. Recall that we use $V_k \triangleq V \times [k]$ to denote the ground set of μ_k. For each $v \in V$ and $i \in [k]$, we denote $v_i \triangleq (v, i) \in V_k$ and $C_v \triangleq \{v_i \mid i \in [k]\}$.

7.1. Spectral independence of k-transformed distribution (proof of Lemma 7.1).

We define a mapping from feasible partial configurations for distribution μ_k to feasible partial configurations for distribution μ. Let $\Lambda \subseteq V_k$ and $\sigma \in \Omega(\mu_{k,\Lambda})$, where $\mu_{k,\Lambda}$ denotes the marginal distribution on Λ projected from μ_k. We define the following subsets of V:

$$
\begin{aligned}
\Lambda^*_{\sigma,-1} &\triangleq \{v \in V \mid \forall i \in [k],\ v_i \in \Lambda \wedge \sigma_{v_i} = -1\}, \\
\Lambda^*_{\sigma,+1} &\triangleq \{v \in V \mid \exists i \in [k] \text{ s.t. } v_i \in \Lambda \wedge \sigma_{v_i} = +1\}, \\
\Lambda^*_\sigma &\triangleq \Lambda^*_{\sigma,-1} \cup \Lambda^*_{\sigma,+1}.
\end{aligned}
\tag{50}
$$

Note that $\Lambda^*_{\sigma,-1}$ and $\Lambda^*_{\sigma,+1}$ are disjoint. Let $\sigma^* \in \{-1, +1\}^{\Lambda^*_\sigma}$ indicate whether each $v \in \Lambda^*_\sigma$ is in $\Lambda^*_{\sigma,+1}$ or $\Lambda^*_{\sigma,-1}$. Specifically,

$$
\forall v \in \Lambda^*_\sigma, \quad \sigma^*_v =
\begin{cases}
-1 & \text{if } v \in \Lambda^*_{\sigma,-1}, \\
+1 & \text{if } v \in \Lambda^*_{\sigma,+1}.
\end{cases}
\tag{51}
$$

In other words, σ^* fixes a $v \in V$ to be -1 if σ fixes all $v_i \in C_v$ to be -1; and σ^* fixes a $v \in V$ to be $+1$ if σ fixes some $v_i \in C_v$ to be $+1$.

Given a $\sigma \in \Omega(\mu_{k,\Lambda})$, we construct the following vector of local fields λ_σ,

$$
\forall v \in V, \quad \lambda_\sigma(v) =
\begin{cases}
1 & \text{if } v \in \Lambda^*_\sigma, \\
\frac{|C_v \setminus \Lambda|}{k} & \text{if } v \notin \Lambda^*_\sigma.
\end{cases}
\tag{52}
$$

By definition, for every $v \in V \setminus \Lambda^*_\sigma$, it holds that $|C_v \setminus \Lambda| > 0$. Hence, λ_σ is a positive vector.

We have the following lemma that bounds the spectral radius of the influence matrix.

Lemma 7.2. *For any integer $k \geq 1$, any $\Lambda \subseteq V_k$ and any $\sigma \in \Omega(\mu_{k,\Lambda})$, it holds that σ^* is a feasible partial configuration with respect to $\nu \triangleq \mu^{(\lambda_\sigma)}$, and*

$$
\rho\left(\Psi^\sigma_{\mu_k}\right) \leq \rho\left(\Psi^{\sigma^*}_\nu\right) + 2,
$$

where $\Psi^\sigma_{\mu_k}$ and $\Psi^{\sigma^*}_\nu$ are the influence matrices defined as in Definition 1.5, and $\rho(\cdot)$ denotes the spectral radius.

Lemma 7.1 immediately holds from Lemma 7.2.

Proof of Lemma 7.1 assuming Lemma 7.2. Fix an integer $k \geq 1$, a subset $\Lambda \subseteq V_k$ and a feasible partial configuration $\sigma \in \Omega(\mu_{k,\Lambda})$. By Lemma 7.2, we have

$$
\rho\left(\Psi^\sigma_{\mu_k}\right) \leq \rho\left(\Psi^{\sigma^*}_\nu\right) + 2, \quad \text{where } \nu \triangleq \mu^{(\lambda_\sigma)}.
$$

Note that $\nu = \mu^{(\lambda_\sigma)}$ and $\lambda_\sigma(v) \in (0, 1)$ for all $v \in V$. Since μ is completely η-spectrally independent,

$$\rho\left(\Psi_\nu^{\sigma^*}\right) \le \eta.$$

This proves the lemma. □

7.2. **Spectrum-preservation of k-transformation (proof of Lemma 7.2).** We first prove that σ^* defined in (51) is feasible with respect to ν as long as $\sigma \in \Omega(\mu_{k,\Lambda})$. By definition of σ^* and μ_k, if $\sigma \in \Omega(\mu_{k,\Lambda})$, then σ^* is feasible with respect to μ. Consequently, σ^* is also feasible with respect to ν, because $\nu = \mu^{(\lambda_\sigma)}$ has the same support as μ for a positive vector λ_σ.

It remains to prove

$$(53) \qquad \rho\left(\Psi_{\mu_k}^\sigma\right) \le \rho\left(\Psi_\nu^{\sigma^*}\right) + 2.$$

For any distribution ν over $\{-1, +1\}^V$, any $H \subseteq V$, and any feasible partial configuration $\sigma_H \in \{-1, +1\}^H$ with respect to ν, let $\mathcal{I}_\nu^{\sigma_H} \in \mathbb{R}_{\ge 0}^{V \times V}$ be a matrix defined as:

$$(54) \qquad \forall u, v \in V, \quad \mathcal{I}_\nu^{\sigma_H}(u, v) \triangleq \max_{i,j \in \Omega\left(v_u^{\sigma_H}\right)} d_{\mathrm{TV}}\left(\nu_v^{\sigma_H, u \leftarrow i}, \nu_v^{\sigma_H, u \leftarrow j}\right).$$

Note that by definition, if $\left|\Omega(v_u^{\sigma_H})\right| = 1$ or $\left|\Omega(v_v^{\sigma_H})\right| = 1$, then $\mathcal{I}_\nu^{\sigma_H}(u, v) = 0$, and for all $u \in V$, $\mathcal{I}_\nu^{\sigma_H}(u, u) = \mathbf{1}[|\Omega(v_v^{\sigma_H})| \ne 1]$.

To prove (53), we define two new matrices $\widetilde{\Psi}_{\mu_k}^\sigma$ and $\widehat{\Psi}_\nu^{\sigma^*}$. Recall that $\sigma \in \{-1, +1\}^\Lambda$ is a partial configuration on $\Lambda \subseteq V_k$. For each variable $v \in V$, we denote by

$$m_v \triangleq |C_v \setminus \Lambda|$$

the number of variables in C_v whose values are not fixed by σ. Without loss of generality, we enumerate the variables in $C_v \setminus \Lambda$ as

$$C_v \setminus \Lambda = \{v_i \mid i \in [m_v]\}, \text{where } [m_v] = \{1, 2, 3, \ldots, m_v\}.$$

The matrix $\widetilde{\Psi}^\sigma \in \mathbb{R}^{(V_k \setminus \Lambda) \times (V_k \setminus \Lambda)}$ is defined by

$$(55) \qquad \forall u, v \in V, i \in [m_u], j \in [m_v], \quad \widetilde{\Psi}^\sigma(u_i, v_j) \triangleq \begin{cases} \frac{2}{m_v} & \text{if } u = v \text{ and } v \notin \Lambda_\sigma^*, \\ \frac{1}{m_v} \mathcal{I}_\nu^{\sigma^*}(u, v) & \text{if } u \ne v \text{ or } v \in \Lambda_\sigma^*. \end{cases}$$

This is well-defined because it only involves those $u, v \in V$ such that $m_v \ge 1$ and $m_u \ge 1$. The following is easy to observe.

Observation 7.3. *For any $u, v \in V$, all the entries $\widetilde{\Psi}^\sigma(u_i, v_j)$ have the same value for $i \in [m_u]$ and $j \in [m_v]$.*

Next, we define the matrix $\widehat{\Psi}^{\sigma^*} \in \mathbb{R}^{(V \setminus \Lambda_\sigma^*) \times (V \setminus \Lambda_\sigma^*)}$, where Λ_σ^* is defined in (50). By Observation 7.3, $\widetilde{\Psi}^\sigma$ can be decomposed into a set of blocks, all the entries in the same block have the same value. The matrix $\widehat{\Psi}^{\sigma^*}$ is defined by compressing blocks in $\widetilde{\Psi}^\sigma$. Formally,

$$(56) \qquad \forall u, v \in V \setminus \Lambda_\sigma^*, \quad \widehat{\Psi}^{\sigma^*}(u, v) \triangleq \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j).$$

The matrix $\widehat{\Psi}^{\sigma^*}$ is well-defined because by the definition of Λ_σ^*, it is straightforward to verify $m_u \ge 1$ and $m_v \ge 1$ for all $u, v \in V \setminus \Lambda_\sigma^*$, thus the u_1 in (56) exists. By Observation 7.3, it holds that for all $u, v \in V \setminus \Lambda_\sigma^*$,

$$\forall i \in [m_u], \quad \widehat{\Psi}^{\sigma^*}(u, v) = \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_i, v_j).$$

The next lemma bounds the relation between spectral radiuses of $\Psi_{\mu_k}^\sigma$, $\Psi_\nu^{\sigma^*}$ and $\widetilde{\Psi}^\sigma$, $\widehat{\Psi}^{\sigma^*}$.

Lemma 7.4. *It holds that $\rho\left(\Psi_{\mu_k}^\sigma\right) \le \rho\left(\widetilde{\Psi}^\sigma\right)$ and $\rho\left(\widehat{\Psi}^{\sigma^*}\right) \le \rho\left(\Psi_\nu^{\sigma^*}\right) + 2$.*

We first use Lemma 7.4 to prove Lemma 7.2, then prove Lemma 7.4.

Proof of Lemma 7.2 assuming Lemma 7.4. Due to Lemma 7.4, it suffices to show that

$$\rho\left(\widetilde{\Psi}^\sigma\right) \le \rho\left(\widehat{\Psi}^{\sigma^*}\right).$$

We will prove that for any eigenvalue $\lambda_{\text{eig}} \in \mathbb{C}$ of $\widetilde{\Psi}^\sigma$, if $\lambda_{\text{eig}} \ne 0$, then λ_{eig} is also an eigenvalue of $\widehat{\Psi}^{\sigma^*}$. Hence, $\rho\left(\widetilde{\Psi}^\sigma\right) \le \rho\left(\widehat{\Psi}^{\sigma^*}\right)$, this proves the lemma.

Fix an eigenvalue $\lambda_{\text{eig}} \ne 0$ of $\widetilde{\Psi}^\sigma$. Let $f : V_k \setminus \Lambda \to \mathbb{C}$ denote the corresponding eigenvector. For any variable $v \in V$, any $i, j \in [m_v]$, by Observation 7.3, it holds that $\widetilde{\Psi}^\sigma(v_i, \cdot) = \widetilde{\Psi}^\sigma(v_j, \cdot)$. We have

$$\lambda_{\text{eig}} f(v_i) = (\widetilde{\Psi}^\sigma f)(v_i) = (\widetilde{\Psi}^\sigma f)(v_j) = \lambda_{\text{eig}} f(v_j).$$

Since $\lambda_{\text{eig}} \ne 0$, we have $f(v_i) = f(v_j)$.

Next, we define a vector $g : V \setminus \Lambda_\sigma^* \to \mathbb{C}$ according to f. For any $v \in V \setminus \Lambda_\sigma^*$, by (50), $m_v \ge 1$. Define

$$g(v) = f(v_1).$$

It holds that

(57) $$\forall v \in V \setminus \Lambda_\sigma^*, i \in [m_v], \quad g(v) = f(v_i).$$

We show that if f is an eigenvector of $\widetilde{\Psi}^\sigma$ with eigenvalue $\lambda_{\text{eig}} \ne 0$, then g is not a zero-vector. This fact can be verified according to the following two cases.

- If there exists $v \notin \Lambda_\sigma^*$ such that $f(v_1) \ne 0$, then g is not a zero-vector.
- Otherwise, for all $v \in V$ and $i \in [m_v]$, if $f(v_i) \ne 0$, then $v \in \Lambda_\sigma^*$. For any $v \in \Lambda_\sigma^*$ and $i \in [m_v]$, by (55), it is straightforward to verify that $\widetilde{\Psi}^\sigma(w, v_i) = 0$ for all $w \in V_k \setminus \Lambda$, thus it holds that $\widetilde{\Psi}^\sigma f = \mathbf{0}$. Since $\lambda_{\text{eig}} \ne 0$, this case cannot occur.

We show that λ_{eig} is also an eigenvalue of $\widehat{\Psi}^{\sigma^*}$ with eigenvector g. We have:

$$\forall u \in V \setminus \Lambda_\sigma^*, \quad \left(\widehat{\Psi}^{\sigma^*} g\right)(u) = \sum_{v \in V \setminus \Lambda_\sigma^*} \widehat{\Psi}^{\sigma^*}(u, v) g(v)$$

$$= \sum_{v \in V \setminus \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) g(v) \qquad \text{(by (56))}$$

(58) $$= \sum_{v \in V \setminus \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j). \qquad \text{(by (57))}$$

We claim that the following equation holds

(59) $$\sum_{v \in V \setminus \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j) = \sum_{v \in V} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j).$$

To verify (59), we consider a $v \in \Lambda_\sigma^*$ with $m_v \ge 1$. Since $u \in V \setminus \Lambda_\sigma^*$, it must hold that $u \ne v$. Since $v \in \Lambda_\sigma^*$, the value of v is fixed by σ^*, thus $|\Omega(v_v^{\sigma^*})| = 1$. By definition in (54), we have

$$\forall v \in \Lambda_\sigma^*, j \in [m_v], \quad \widetilde{\Psi}^\sigma(u_1, v_j) = \frac{1}{m_v} I_v^{\sigma^*}(u, v) = 0.$$

Hence

$$\sum_{v \in V} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j) = \sum_{v \in V \setminus \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j) + \sum_{v \in \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j)$$

$$= \sum_{v \in V \setminus \Lambda_\sigma^*} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j).$$

This proves (59). Combining (59) and (58), we have

$$\forall u \in V \setminus \Lambda_\sigma^*, \quad \left(\widehat{\Psi}^{\sigma^*} g\right)(u) = \sum_{v \in V} \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) f(v_j) = \left(\widetilde{\Psi}^\sigma f\right)(u_1) = \lambda_{\text{eig}} f(u_1) = \lambda_{\text{eig}} g(u).$$

Since g is a non-zero vector, λ_{eig} is also an eigenvalue of $\widehat{\Psi}^{\sigma^*}$. □

7.3. Spectral radius bounds for the intermediate matrices (proof of Lemma 7.4).

Let $A, B \in \mathbb{R}^{n \times n}$, if for all $i, j \in [n]$, $A(i, j) \le B(i, j)$, then we denote $A \le B$. We need the following proposition in linear algebra.

Proposition 7.5 ([? , Corollary 8.1.19]). *Let $A, B \in \mathbb{R}_{\ge 0}^{n \times n}$ be two nonnegative matrices. If $A \le B$, then it holds that $\rho(A) \le \rho(B)$.*

We first prove that

$$\tag{60} \rho\left(\widehat{\Psi}^{\sigma^*}\right) \le \rho\left(\Psi_v^{\sigma^*}\right) + 2$$

By the definition of $\widehat{\Psi}^{\sigma^*}$ in (56), for any $u, v \in V \setminus \Lambda_\sigma^*$ satisfying $u \ne v$, it holds that

$$\widehat{\Psi}^{\sigma^*}(u, v) = \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(u_1, v_j) = \sum_{j \in [m_v]} \frac{\mathcal{I}_v^{\sigma^*}(u, v)}{m_v} = \mathcal{I}_v^{\sigma^*}(u, v) = \Psi_v^{\sigma^*}(u, v),$$

where the last equation holds due to Definition 1.5 and the fact $u \ne v$. For any $v \in V \setminus \Lambda_\sigma^*$,

$$\widehat{\Psi}^{\sigma^*}(v, v) = \sum_{j \in [m_v]} \widetilde{\Psi}^\sigma(v_1, v_j) = \sum_{j \in [m_v]} \frac{2}{m_v} \le 2.$$

Combining above inequalities together, we have

$$\widehat{\Psi}^{\sigma^*} \le \Psi_v^{\sigma^*} + 2I,$$

where I is the identity matrix. By Proposition 7.5, we have

$$\rho\left(\widehat{\Psi}^{\sigma^*}\right) \le \rho\left(\Psi_v^{\sigma^*} + 2I\right) \le \rho\left(\Psi_v^{\sigma^*}\right) + 2.$$

This proves (60).

We next prove that

$$\tag{61} \rho\left(\Psi_{\mu_k}^\sigma\right) \le \rho\left(\widetilde{\Psi}^\sigma\right).$$

By Proposition 7.5, it suffices to show that $\Psi_{\mu_k}^\sigma \le \widetilde{\Psi}^\sigma$. By Definition 1.5, for all $w \in V_k \setminus \Lambda$, it holds that $\Psi_{\mu_k}^\sigma(w, w) = 0$, and for all $w, w' \in V_k \setminus \Lambda$ with $w \ne w'$, it holds that $\Psi_{\mu_k}^\sigma(w, w') = \mathcal{I}_{\mu_k}^\sigma(w, w')$. By the definition of $\widetilde{\Psi}^\sigma$ in (55), to prove Lemma 7.4, it only remains to verify the following facts:

(62) $\forall u, v \in V$ with $v \in \Lambda_\sigma^*, \forall i \in [m_u], j \in [m_v]: \quad \Psi_{\mu_k}^\sigma(u_i, v_j) = \widetilde{\Psi}^\sigma(u_i, v_j) = 0$;

(63) $\forall u, v \in V$ with $u \ne v, \forall i \in [m_u], j \in [m_v]: \quad \Psi_{\mu_k}^\sigma(u_i, v_j) = \mathcal{I}_{\mu_k}^\sigma(u_i, v_j) \le \frac{1}{m_v} \mathcal{I}_v^{\sigma^*}(u, v) = \widetilde{\Psi}^\sigma(u_i, v_j)$;

(64) $\forall u \in V, i, j \in [m_u]$ with $v \notin \Lambda_\sigma^*$ and $i \ne j: \quad \Psi_{\mu_k}^\sigma(u_i, u_j) = \mathcal{I}_{\mu_k}^\sigma(u_i, u_j) \le \frac{2}{m_u} = \widetilde{\Psi}^\sigma(u_i, u_j)$.

We first verify (62), which holds trivially because $|\Omega(\mu_{k, v_j}^\sigma)| = |\Omega(v_v^{\sigma^*})| = 1$.

The rest of this section is dedicated to verifying the facts given in (63) and (64). Without loss of generality, when proving (63) and (64), we assume that

$$\Omega(\mu_{k, u_i}^\sigma) = \{-1, +1\} \text{ and } \Omega(\mu_{k, v_j}^\sigma) = \{-1, +1\},$$
$$\Omega(\mu_{k, u_i}^\sigma) = \{-1, +1\} \text{ and } \Omega(\mu_{k, u_j}^\sigma) = \{-1, +1\}.$$

Otherwise, if $\Omega(\mu^\sigma_{k,u_i}) \neq \{-1,+1\}$ or $\Omega(\mu^\sigma_{k,v_j}) \neq \{-1,+1\}$, then it must hold that $|\Omega(\mu^\sigma_{k,u_i})| = 1$ or $|\Omega(\mu^\sigma_{k,v_j})| = 1$, thus $\mathcal{I}^\sigma_{\mu_k}(u_i, v_j) = 0$ and (63) holds trivially; if $\Omega(\mu^\sigma_{k,u_i}) \neq \{-1,+1\}$ or $\Omega(\mu^\sigma_{k,u_j}) \neq \{-1,+1\}$, then it must hold that $|\Omega(\mu^\sigma_{k,u_i})| = 1$ or $|\Omega(\mu^\sigma_{k,u_j})| = 1$, thus $\mathcal{I}^\sigma_{\mu_k}(u_i, u_j) = 0$ and (64) holds trivially. By our assumptions, we can conclude that

$$u \notin \Lambda^*_\sigma \text{ and } v \notin \Lambda^*_\sigma.$$

Because if $u \in \Lambda^*_\sigma$, then it holds that $|\Omega(\mu^\sigma_{k,u_i})| = 1$; and if $v \in \Lambda^*_\sigma$, then it holds that $|\Omega(\mu^\sigma_{k,v_j})| = 1$.

We first prove (63). Note that $\mathcal{I}^\sigma_{\mu_k}(u_i, v_j) = d_{\text{TV}}\left(\mu^{\sigma, u_i \leftarrow +1}_{k,v_j}, \mu^{\sigma, u_i \leftarrow -1}_{k,v_j}\right)$. By definition, we have

$$\mu^{\sigma, u_i \leftarrow +1}_{k,v_j}(+1) = \frac{\mathbf{Pr}_{Y \sim \mu_k}\left[Y_{v_j} = +1 \wedge Y_{u_i} = +1 \wedge Y_\Lambda = \sigma\right]}{\mathbf{Pr}_{Y \sim \mu_k}\left[Y_{u_i} = +1 \wedge Y_\Lambda = \sigma\right]}.$$

Recall $\Lambda^*_{\sigma,+1}$ is defined in (50). Let $N = |\Lambda^*_{\sigma,+1}|$. By the definition of distribution μ_k, we have

$$\mathbf{Pr}_{Y \sim \mu_k}\left[Y_{v_j} = +1 \wedge Y_{u_i} = +1 \wedge Y_\Lambda = \sigma\right]$$

(65)
$$= \sum_{X \in \Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w \in V \setminus \Lambda^*_\sigma : X_w = +1} \frac{m_w}{k}\right)\left(\frac{\mathbf{1}\left[X_v = +1\right]}{m_v}\right)\left(\frac{\mathbf{1}\left[X_u = +1\right]}{m_u}\right).$$

Recall that to draw a random sample $Y \sim \mu_k$, one needs to draw a random sample $X \sim \mu$, then transform X to Y according to the rules in Definition 2.6. Equation (65) holds due to the following arguments:

- If $Y_\Lambda = \sigma$, by definitions in (50) and (51), it must have $X_{\Lambda^*_\sigma} = \sigma^*$ and for every $w \in \Lambda^*_{\sigma,+1}$, a particular variable $w_\ell \in C_w$ (determined by σ) is picked and is assigned with value $+1$. The probability of such event is $(\frac{1}{k})^N$. Furthermore, for every $w \in V \setminus \Lambda^*_\sigma$, σ requires that $k - m_w$ variables in C_w must take value -1. If $X_w = +1$, a variable among the remaining m_w variables should be picked and assigned with value $+1$. Such event occurs with probability $\prod_{w \in V \setminus \Lambda^*_\sigma : X_w = +1} \frac{m_w}{k}$.
- If $Y_{v_j} = +1$, then $X_v = +1$, and v_j must be picked and is assigned with value $+1$. Since $v \notin \Lambda^*_\sigma$, conditional on previous events, this event holds with probability $\frac{1}{m_v}$.
- If $Y_{u_i} = +1$, then $X_u = +1$, and u_i must be picked and is assigned with value $+1$. Since $u \notin \Lambda^*_\sigma$, conditional on previous events, this event holds with probability $\frac{1}{m_u}$.

Similarly, we have

$$\mathbf{Pr}_{Y \sim \mu_k}\left[Y_{u_i} = +1 \wedge Y_\Lambda = \sigma\right] = \sum_{X \in \Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w \in V \setminus \Lambda^*_\sigma : X_w = +1} \frac{m_w}{k}\right)\left(\frac{\mathbf{1}\left[X_u = +1\right]}{m_u}\right).$$

Recall $\nu = \mu^{(\lambda_\sigma)}$, where the local fields λ_σ are constructed in (52). We have

$$\mu^{\sigma, u_i \leftarrow +1}_{k,v_j}(+1) = \frac{\sum_{X \in \Omega(\mu)} \nu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{\mathbf{1}[X_v = +1]}{m_v}\right)\left(\frac{\mathbf{1}[X_u = +1]}{m_u}\right)}{\sum_{X \in \Omega(\mu)} \nu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{\mathbf{1}[X_u = +1]}{m_u}\right)}$$

$$= \frac{1}{m_v}\frac{\mathbf{Pr}_{X \sim \nu}\left[X_{\Lambda^*_\sigma} = \sigma^* \wedge X_v = +1 \wedge X_u = +1\right]}{\mathbf{Pr}_{X \sim \nu}\left[X_{\Lambda^*_\sigma} = \sigma^* \wedge X_u = +1\right]}$$

(66)
$$= \frac{\nu^{\sigma^*, u \leftarrow +1}_v(+1)}{m_v}.$$

Note that since $+1 \in \Omega(\mu^\sigma_{k,u_i})$, we have $+1 \in \Omega(\nu^{\sigma^*}_u)$, and thus $\nu^{\sigma^*, u \leftarrow +1}_v(+1)$ is well defined.

Next, we calculate

$$\mu_{k,v_j}^{\sigma,u_i\leftarrow -1}(+1) = \frac{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{v_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}.$$

By the definition of distribution μ_k, we have

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{v_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]$$

$$= \sum_{X\in\Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{\substack{w\in V\setminus\Lambda_\sigma^* \\ X_w=+1}} \frac{m_w}{k}\right)\left(\frac{\mathbf{1}\left[X_v = +1\right]}{m_v}\right)\left(\mathbf{1}\left[X_u = +1\right]\frac{m_u - 1}{m_u} + \mathbf{1}\left[X_u = -1\right]\right).$$

Compared with (65), the only difference is calculating the probability of event $Y_{u_i} = -1$ conditional on other events. Note that $u \notin \Lambda_\sigma^*$. If $X_u = -1$, then it must hold that $Y_{u_i} = -1$; and if otherwise $X_u = +1$, a variable among all m_u unfixed variables in C_u is picked, and such variable cannot be u_i, which occurs with probability $\frac{m_u - 1}{m_u}$. Similarly, we have

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]$$

$$= \sum_{X\in\Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w\in V\setminus\Lambda_\sigma^*:X_w=+1} \frac{m_w}{k}\right)\left(\mathbf{1}\left[X_u = +1\right]\frac{m_u - 1}{m_u} + \mathbf{1}\left[X_u = -1\right]\right).$$

We define a vector of local fields $\boldsymbol{\lambda}_\sigma'$ as

$$\forall w \in V, \quad \lambda_\sigma'(w) = \begin{cases} \frac{m_u - 1}{k} & \text{if } w = u, \\ \frac{m_w}{k} & \text{if } w \in V \setminus (\{u\} \cup \Lambda_\sigma^*), \\ 1 & \text{if } w \in \Lambda_\sigma^*. \end{cases}$$

Note that $\boldsymbol{\lambda}_\sigma'$ differs from $\boldsymbol{\lambda}_\sigma$ only at u, where $\lambda_\sigma'(u) = \frac{m_u - 1}{k}$ and $\lambda_\sigma(u) = \frac{m_u}{k}$. Since $u, v \in V \setminus \Lambda_\sigma^*$, we have

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{v_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right] = \sum_{X\in\Omega(\mu)} \frac{\mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*} = \sigma^*\right]}{k^N}\left(\prod_{\substack{w\in V\setminus\Lambda_\sigma^* \\ X_w=+1}} \lambda_\sigma'(w)\right)\left(\frac{\mathbf{1}\left[X_v = +1\right]}{m_v}\right),$$

and

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right] = \sum_{X\in\Omega(\mu)} \frac{\mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*} = \sigma^*\right]}{k^N}\left(\prod_{\substack{w\in V\setminus\Lambda_\sigma^* \\ X_w=+1}} \lambda_\sigma'(w)\right).$$

We define a new distribution π by imposing the local fields $\boldsymbol{\lambda}_\sigma'$ to μ:

$$\pi = \mu^{(\boldsymbol{\lambda}_\sigma')}.$$

If $m_u > 1$, then $\boldsymbol{\lambda}_\sigma'$ is a positive vector, the distribution π is well-defined. If $m_u = 1$, then $\lambda_\sigma'(u) = 0$ and $\lambda_\sigma'(w) > 0$ for all $w \neq u$. In this case, variable u can only take value -1, and the distribution π is well-defined as long as $\mu_u(-1) > 0$. Note that $\Omega(\mu_{k,u_i}^\sigma) = \{-1, +1\}$. Conditional on σ, u_i takes value -1 with positive probability in distribution μ_k. Since $m_u = 1$, with a positive probability, all variables in C_u take value -1. By the definition of distribution μ_k, we have

(67)
$$m_u = 1 \quad \Longrightarrow \quad \mu_u^{\sigma^*}(-1) > 0.$$

Hence $\mu_u(-1) > 0$ and the distribution π is always well-defined.

34

We have

$$\mu_{k,v_j}^{\sigma,u_i\leftarrow-1}(+1) = \frac{\sum_{X\in\Omega(\mu)}\mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*}=\sigma^*\right]\left(\frac{\mathbf{1}[X_v=+1]}{m_v}\right)\left(\prod_{w\in V\setminus\Lambda_\sigma^*:X_w=+1}\lambda_\sigma'(w)\right)}{\sum_{X\in\Omega(\mu)}\mu(X)\mathbf{1}\left[X_{\Lambda_\sigma^*}=\sigma^*\right]\left(\prod_{w\in V\setminus\Lambda_\sigma^*:X_w=+1}\lambda_\sigma'(w)\right)}$$

$$= \frac{\sum_{X\in\Omega(\mu)}\pi(X)\mathbf{1}\left[X_{\Lambda_\sigma^*}=\sigma^*\right]\left(\frac{\mathbf{1}[X_v=+1]}{m_v}\right)}{\sum_{X\in\Omega(\mu)}\pi(X)\mathbf{1}\left[X_{\Lambda_\sigma^*}=\sigma^*\right]}$$

$$= \frac{1}{m_v}\frac{\mathbf{Pr}_{X\sim\pi}\left[X_{\Lambda_\sigma^*}=\sigma^*\wedge X_v=+1\right]}{\mathbf{Pr}_{X\sim\pi}\left[X_{\Lambda_\sigma^*}=\sigma^*\right]}$$

(68)
$$= \frac{\pi_v^{\sigma^*}(+1)}{m_v}.$$

Combining (66) and (68), we have

$$\mathcal{I}_{\mu_k}^\sigma(u_i,v_j) = d_{\mathrm{TV}}\left(\mu_{k,v_j}^{\sigma,u_i\leftarrow+1},\mu_{v_j}^{\sigma,u_i\leftarrow-1}\right)$$

$$= \left|\mu_{k,v_j}^{\sigma,u_i\leftarrow+1}(+1)-\mu_{v_j}^{\sigma,u_i\leftarrow-1}(+1)\right|$$

$$= \frac{1}{m_v}\left|v_v^{\sigma^*,u\leftarrow+1}(+1)-\pi_v^{\sigma^*}(+1)\right|$$

(69)
$$= \frac{1}{m_v}d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow+1},\pi_v^{\sigma^*}\right).$$

Finally, we bound $d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow+1},\pi_v^{\sigma^*}\right)$. We construct the following coupling between $v_v^{\sigma^*,u\leftarrow+1}$ and $\pi_v^{\sigma^*}$:

- sample a random value $c\in\{-1,+1\}$ according to the distribution $\pi_u^{\sigma^*}$;
- sample c_v, c_v' jointly according to the optimal coupling between $v_v^{\sigma^*,u\leftarrow+1}$ and $\pi_v^{\sigma^*,u\leftarrow c}$.

Recall that $v = \mu^{(\lambda_\sigma)}$ and $\pi = \mu^{(\lambda_\sigma)}$. We claim that

(70)
$$+1\in\Omega(v_u^{\sigma^*})\quad\text{and}\quad\Omega(\pi_u^{\sigma^*})\subseteq\Omega(v_u^{\sigma^*}).$$

Assume that (70) is correct. Recall that λ_σ' differs from λ_σ only at u, where $\lambda_\sigma'(u)=\frac{m_u-1}{k}$ and $\lambda_\sigma(u)=\frac{m_u}{k}$. It is straightforward to verify that

$$\forall c\in\Omega(\pi_u^{\sigma^*}),\quad \pi_v^{\sigma^*,u\leftarrow c}=v_v^{\sigma^*,u\leftarrow c}.$$

By the coupling inequality, we have

$$d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow+1},\pi_v^{\sigma^*}\right) \le \mathbf{Pr}\left[c_v\neq c_v'\right]$$

$$= \sum_{c\in\Omega(\pi_u^{\sigma^*})}\pi_u^{\sigma^*}(c)d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow+1},\pi_v^{\sigma^*,u\leftarrow c}\right)$$

$$= \sum_{c\in\Omega(\pi_u^{\sigma^*})}\pi_u^{\sigma^*}(c)d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow+1},v_v^{\sigma^*,u\leftarrow c}\right)$$

$$\le \max_{i,j\in\Omega(v_u^{\sigma^*})}d_{\mathrm{TV}}\left(v_v^{\sigma^*,u\leftarrow i},v_v^{\sigma^*,u\leftarrow j}\right)$$

(71)
$$= \mathcal{I}_v^{\sigma^*}(u,v).$$

Combining (69) and (71) proves (63).

35

We now verify (70). Since $\boldsymbol{\lambda}_\sigma$ is a positive vector, then ν and μ have the same support. Note that we assume $\Omega(\mu^\sigma_{k,u_i}) = \{-1,+1\}$. By the definition of μ_k and σ^*, it holds that $+1 \in \Omega(\mu^{\sigma^*}_u)$, hence $+1 \in \Omega(\nu^{\sigma^*}_u)$. We now show that $\Omega(\pi^{\sigma^*}_u) \subseteq \Omega(\nu^{\sigma^*}_u)$. Note that $\lambda'_\sigma(u) = \frac{m_u-1}{k}$ and $\lambda_\sigma(u) = \frac{m_u}{k}$. There are two cases:

- If $m_u > 1$, then $\boldsymbol{\lambda}_\sigma$ and $\boldsymbol{\lambda}'_\sigma$ are both positive, ν and π have the same support. Thus, $\Omega(\pi^{\sigma^*}_u) = \Omega(\nu^{\sigma^*}_u)$.
- If $m_u = 1$, in this case $\Omega(\pi^{\sigma^*}_u) = \{-1\}$, by (67), $-1 \in \Omega(\mu^{\sigma^*}_u)$. Since $\boldsymbol{\lambda}_\sigma$ is a positive vector, μ and ν have the same support, thus $-1 \in \Omega(\nu^{\sigma^*}_u)$.

Next, we prove (64). Recall that without loss of generality $\Omega(\mu^\sigma_{k,u_i}) = \{-1,+1\}$. We have

$$\mathcal{I}^\sigma_{\mu_k}(u_i, u_j) = d_{\mathrm{TV}}\left(\mu^{\sigma,u_i\leftarrow+1}_{k,u_j}, \mu^{\sigma,u_i\leftarrow-1}_{k,u_j}\right) = \mu^{\sigma,u_i\leftarrow-1}_{k,u_j}(+1),$$

which is due to that $\mu^{\sigma,u_i\leftarrow+1}_{k,u_j}(+1) = 0$. By definition,

$$\mu^{\sigma,u_i\leftarrow-1}_{k,u_j}(+1) = \frac{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}.$$

Recall $N = |\Lambda^*_{\sigma,+1}|$. By the definition of distribution μ_k, we have

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]$$

$$= \sum_{X\in\Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w\in V\backslash\Lambda^*_\sigma : X_w=+1} \frac{m_w}{k}\right) \frac{\mathbf{1}\left[X_u = +1\right]}{m_u},$$

Compared with (65), the only difference is calculating the probability of event $Y_{u_i} = -1 \wedge Y_{u_j} = +1$ conditional on other events. Note that $u \notin \Lambda^*_\sigma$. If $Y_{u_i} = -1 \wedge Y_{u_j} = +1$, then $X_u = +1$, and we pick variable u_j to set its value to $+1$. Conditional on other events, this event occurs with probability $\frac{1}{m_u}$.

We assume $\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right] > 0$, since if otherwise $\mathcal{I}^\sigma_{\mu_k}(u_i, u_j) = \mu^{\sigma,u_i\leftarrow-1}_{k,u_j}(+1) = 0$ and (64) holds trivially. Similarly,

$$\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]$$

$$= \sum_{X\in\Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w\in V\backslash\Lambda^*_\sigma : X_w=+1} \frac{m_w}{k}\right)\left(\mathbf{1}\left[X_u = +1\right]\frac{m_u-1}{m_u} + \mathbf{1}\left[X_u = -1\right]\right)$$

$$\geq \sum_{X\in\Omega(\mu)} \mu(X)\mathbf{1}\left[X_{\Lambda^*_\sigma} = \sigma^*\right]\left(\frac{1}{k}\right)^N \left(\prod_{w\in V\backslash\Lambda^*_\sigma : X_w=+1} \frac{m_w}{k}\right)\left(\mathbf{1}\left[X_u = +1\right]\frac{m_u-1}{m_u}\right)$$

$$= (m_u - 1)\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right].$$

In (64), we assume $i \neq j$, which implies $m_u \geq 2$. We have

$$\mathcal{I}^\sigma_{\mu_k}(u_i, u_j) = \frac{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_j} = +1 \wedge Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}{\mathbf{Pr}_{Y\sim\mu_k}\left[Y_{u_i} = -1 \wedge Y_\Lambda = \sigma\right]}$$

$$\leq \frac{1}{m_u - 1}$$

$$\leq \frac{2}{m_u},$$

which proves (64). Together with the above proof of (63), this proves Lemma 7.4, which concludes our proof of Lemma 2.8.

8. Spectral Gaps of Two-Spin Systems

In this section, we prove Theorem 2.10 and its slightly strengthened variants for the hardcore and Ising models, which respectively imply Theorem 1.1, Theorem 1.3 and Theorem 1.4.

Let $\mathcal{I} = (V, E, \lambda, \beta, \gamma)$ denote a two-spin system on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ specified by parameters (β, γ, λ). Let μ denote the Gibbs distribution induced by $\mathcal{I}$. We assume that the system is anti-ferromagnetic, that is, $0 \leq \beta \leq \gamma$, $\beta\gamma < 1$, and $\gamma, \lambda > 0$. Let $\delta \in (0, 1)$ be a positive gap. We assume that (β, γ, λ) is up-to-Δ unique with gap δ (Definition 3.2).

Given any $\chi \in \{-1, +1\}^V$, recall the distribution $\nu = \mathrm{flip}(\mu, \chi)$ obtained by flipping μ according to χ:

$$\forall \sigma \in \{-1, +1\}^V, \quad \nu(\sigma) \triangleq \mu(\sigma \odot \chi),$$

where $\sigma \odot \chi \in \{-1, +1\}^V$ satisfying $(\sigma \odot \chi)_v = \sigma_v \chi_v$ for all $v \in V$. Obviously, $\mu = \mathrm{flip}(\mathrm{flip}(\mu, \chi), \chi)$.

Given any $\chi \in \{-1, +1\}^V$, for any vector $\boldsymbol{\theta} \in (0, 1]^V$ or scalar $\theta \in (0, 1]$, we write:

$$\tag{72} \boldsymbol{\theta}^{\boldsymbol{\chi}} \triangleq (\theta_v^{\chi_v})_{v \in V} \quad \text{and} \quad \theta^{\boldsymbol{\chi}} \triangleq (\theta^{\chi_v})_{v \in V}.$$

The following fact for flipping follows easily from the factorization of local fields.

Fact 8.1. *For any $\boldsymbol{\theta} \in (0, 1]^V$ and $\chi \in \{-1, +1\}^V$, $\nu^{(\boldsymbol{\theta})} = \mathrm{flip}(\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}, \chi)$ where $\nu = \mathrm{flip}(\mu, \chi)$.*

Proof. For all $\boldsymbol{\theta} \in (0, 1]^V$ and $\sigma, \tau \in \Omega(\nu)$,

$$\frac{\nu^{(\boldsymbol{\theta})}(\sigma)}{\nu^{(\boldsymbol{\theta})}(\tau)} = \frac{\nu(\sigma) \prod_{u \in \sigma^{-1}(+1)} \theta_u}{\nu(\tau) \prod_{v \in \tau^{-1}(+1)} \theta_v} = \frac{\mu(\sigma \odot \chi) \prod_{u \in \sigma^{-1}(+1)} \theta_u}{\mu(\tau \odot \chi) \prod_{v \in \tau^{-1}(+1)} \theta_v} \overset{(*)}{=} \frac{\mu(\sigma \odot \chi) \prod_{u \in (\sigma \odot \chi)^{-1}(+1)} \theta_u^{\chi_u}}{\mu(\tau \odot \chi) \prod_{v \in (\tau \odot \chi)^{-1}(+1)} \theta_v^{\chi_v}} = \frac{\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}(\sigma \odot \chi)}{\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}(\tau \odot \chi)},$$

where the equation $(*)$ holds because

$$\frac{\prod_{u \in (\sigma \odot \chi)^{-1}(+1)} \theta_u^{\chi_u}}{\prod_{v \in (\tau \odot \chi)^{-1}(+1)} \theta_v^{\chi_v}} = \frac{\prod_{u \in \sigma^{-1}(+1) \cap \chi^{-1}(+1)} \theta_u \cdot \prod_{v \in \sigma^{-1}(-1) \cap \chi^{-1}(-1)} \frac{1}{\theta_v} \cdot \prod_{w \in \chi^{-1}(-1)} \theta_w}{\prod_{u \in \tau^{-1}(+1) \cap \chi^{-1}(+1)} \theta_u \cdot \prod_{v \in \tau^{-1}(-1) \cap \chi^{-1}(-1)} \frac{1}{\theta_v} \cdot \prod_{w \in \chi^{-1}(-1)} \theta_w}$$

$$= \frac{\prod_{u \in \sigma^{-1}(+1) \cap \chi^{-1}(+1)} \theta_u \cdot \prod_{v \in \sigma^{-1}(+1) \cap \chi^{-1}(-1)} \theta_v}{\prod_{u \in \tau^{-1}(+1) \cap \chi^{-1}(+1)} \theta_u \cdot \prod_{v \in \tau^{-1}(+1) \cap \chi^{-1}(-1)} \theta_v} = \frac{\prod_{u \in \sigma^{-1}(+1)} \theta_u}{\prod_{v \in \tau^{-1}(+1)} \theta_v}.$$

The above equation implies $\nu^{(\boldsymbol{\theta})}(\sigma) = \mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}(\sigma \odot \chi)$, i.e. $\nu^{(\boldsymbol{\theta})} = \mathrm{flip}(\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}, \chi)$. □

We generalize the notion of complete spectral independence (Definition 1.8) to different directions.

Definition 8.2 (complete SI in a direction). Let $\chi \in \{-1, +1\}^V$. A distribution μ over $\{-1, +1\}^V$ is said to be *completely η-spectrally independent in direction χ* if $\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}$ is η-spectrally independent for all $\boldsymbol{\theta} \in (0, 1]^V$.

The complete η-spectral independence defined in Definition 1.8 is the special case of the above definition with direction $\chi = \mathbf{1}$, where $\mathbf{1} = (+1)_{v \in V}$ denotes the all-$(+1)$ vector. The following observation follows from Fact 8.1 and the fact that flipping defines isomorphisms for measures and dynamics.

Observation 8.3. *Let $\nu = \mathrm{flip}(\mu, \chi)$ for distribution μ over $\{-1, +1\}^V$ and $\chi \in \{-1, +1\}^V$.*

- *ν is completely η-spectrally independent if and only if μ is completely η-spectrally independent in direction χ;*
- *For any $\boldsymbol{\theta} \in (0, 1]^V$, it holds that $\lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\mu^{(\boldsymbol{\theta}^{\boldsymbol{\chi}})}) = \lambda_{\mathrm{min\text{-}gap}}^{\mathrm{GD}}(\nu^{(\boldsymbol{\theta})})$.*

The good direction $\chi \in \{-1, +1\}^V$ is defined as follows.

$$\tag{73} \forall v \in V, \quad \chi_v = \mathrm{sgn}(\Delta_v, \lambda) \triangleq \begin{cases} +1 & \lambda \leq \left(\frac{\gamma}{\beta}\right)^{\Delta_v/2}, \\ -1 & \text{otherwise}, \end{cases}$$

where Δ_v denotes the degree of v in G and we assume that $\frac{\gamma}{0} = +\infty$ for $\gamma > 0$.

With Observation 8.3, Theorem 2.10 can be equivalently stated as follows.

Lemma 8.4. *For all $\delta \in (0, 1)$, for every anti-ferromagnetic two-spin system on n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$ that is up-to-Δ unique with gap δ, for the $\chi \in \{-1, +1\}^V$ in (73),*

(1) the Gibbs distribution μ is completely $\frac{144}{\delta}$-spectrally independent in direction χ;

(2) for $\theta = \frac{\delta^2}{64}$ and $C = \frac{8}{\delta}$, it holds that $\lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(\theta\chi)}\right) \geq \frac{1}{Cn}$.

The rest of Section 8 is organized as folows: In Section 8.1 we prove the first part of Lemma 8.4, and in Section 8.2 we prove the second part of Lemma 8.4.

8.1. Complete spectral independence up to uniqueness (in the right direction).

Let $\mathcal{I} = (V, E, \beta, \gamma, \lambda)$ denote the anti-ferromagnetic two-spin system specified by (β, γ, λ) on a n-vertex graph $G = (V, E)$ with maximum degree $\Delta = \Delta_G \geq 3$. Let $\theta \in (0, 1]^V$ be arbitrary and $\chi \in \{-1, +1\}^V$ as defined in (73). Let μ denote the Gibbs distribution of $\mathcal{I}$ and $\pi = \mu^{(\theta\chi)}$. Note that $\Omega(\pi) = \Omega(\mu)$ since θ is positive.

We denote by $\mathcal{I}_\pi = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ the two-spin system that is the same as $\mathcal{I}$ except that in $\mathcal{I}_\pi$ each vertex $v \in V$ is associated with the local field

$$(74) \qquad\qquad \lambda_v = \lambda\theta_v^{\chi_v}.$$

It is easy to see that $\pi = \mu^{(\theta\chi)}$ is the Gibbs distribution induced by $\mathcal{I}_\pi$, more specifically:

$$\pi(\sigma) \propto \mu(\sigma) \prod_{v \in V:\sigma(v)=+1} \theta_v^{\chi_v} \propto \beta^{m_{-1}(\sigma)}\gamma^{m_{+1}(\sigma)} \prod_{v \in V:\sigma(v)=+1} \lambda_v,$$

where $m_i(\sigma) \triangleq |\{(u, v) \in E \mid \sigma_u = \sigma_v = i\}|$ for $i \in \{-1, +1\}$. It is then sufficient to verify the $\frac{144}{\delta}$-spectral independence of π for arbitrary $\theta \in (0, 1]^V$.

For every $v \in V$, let Δ_v denote the degree of v in $G = (V, E)$ and $d_v \triangleq \Delta_v - 1$. We have $\Delta = \max_{v \in V} \Delta_v$.

Lemma 8.5. *If (β, γ, λ) is up-to-Δ unique with gap δ, then for every $v \in V$, $(\beta, \gamma, \lambda_v)$ is d_v-unique with gap δ.*

The above lemma is implied by the following proposition from [?].

Proposition 8.6 ([?]). *If $\beta = 0$, then the following holds for all integers $d \geq 1$:*

- *$(0, \gamma, \lambda)$ is d-unique with gap δ iff $\lambda \leq \lambda_{c,\delta}(d) = \frac{(1-\delta)d^d\gamma^{d+1}}{(d-1+\delta)^{d+1}}$.*

Assume $\beta > 0$. Let $\overline{\Delta} \triangleq \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$. The followings hold for all integers $d \geq 1$:

- *If $d < (1 - \delta)\overline{\Delta}$, then (β, γ, λ) is d-unique with gap δ for all $\lambda > 0$.*
- *If $d \geq (1 - \delta)\overline{\Delta}$, then $f_d(x) = 1 - \delta$ has two nonnegative roots $x_1(d)$ and $x_2(d)$ such that*

$$x_1(d) = \frac{\zeta_\delta(d) - \sqrt{\zeta_\delta(d)^2 - 4(1 - \delta)^2\beta\gamma}}{2(1 - \delta)\beta} \quad \text{and} \quad x_2(d) = \frac{\zeta_\delta(d) + \sqrt{\zeta_\delta(d)^2 - 4(1 - \delta)^2\beta\gamma}}{2(1 - \delta)\beta},$$

where $\zeta_\delta(d) \triangleq d(1 - \beta\gamma) - (1 - \delta)(1 + \beta\gamma) > 0$.

- *For $d \geq (1 - \delta)\overline{\Delta}$ and $i \in \{1, 2\}$ let*

$$\lambda_i(d) = x_i(d)\left(\frac{x_i(d) + \gamma}{\beta x_i(d) + 1}\right)^d.$$

It holds that $\lambda_1(d)\lambda_2(d) = \left(\frac{\gamma}{\beta}\right)^{d+1}$. And (β, γ, λ) is d-unique with gap δ iff $\lambda \in (0, \lambda_1(d)] \cup [\lambda_2(d), +\infty)$.

The proposition is slightly refined from the one proved in [?] by taking the gap δ into consideration. A proof of Proposition 8.6 is included in Appendix B.1 for completeness.

Proof of Lemma 8.5. Fix an arbitrary $v \in V$. We consider the following two cases.

Assume $\beta = 0$. Note that $(0, \gamma, \lambda)$ is d_v-unique with gap δ. By Proposition 8.6, $\lambda \leq \lambda_{c,\delta}(d_v)$. According to (73), when $\beta = 0$, $\chi_v = +1$ regardless of Δ_v. Hence, $\lambda_v = \lambda\theta_v \leq \lambda \leq \lambda_{c,\delta}(d_v)$. By Proposition 8.6, $(0, \gamma, \lambda_v)$ is d_v-unique with gap δ.

Assume $\beta > 0$. We may further assume $d_v \geq (1 - \delta)\overline{\Delta}$. Otherwise, by Proposition 8.6, $(\beta, \gamma, \lambda_v)$ is d_v-unique. Since (β, γ, λ) is d_v-unique with gap δ, by Proposition 8.6, it holds that $\lambda \in (0, \lambda_1(d_v)] \cup [\lambda_2(d_v), +\infty)$. We further consider the following two sub-cases.

- If $\lambda \leq \lambda_1(d_v)$, since $\lambda_1(d_v)\lambda_2(d_v) = (\gamma/\beta)^{d_v+1} = (\gamma/\beta)^{\Delta_v}$, we have $\lambda \leq \lambda_1(d_v) \leq (\gamma/\beta)^{\frac{\Delta_v}{2}}$. By (73), we have $\lambda_v = \lambda\theta_v \leq \lambda \leq \lambda_1(d_v)$.

- If $\lambda > \lambda_1(d_v)$, since (β, λ, γ) is d_v-unique, we have $\lambda \geq \lambda_2(d_v) \geq (\gamma/\beta)^{\frac{\Delta_v}{2}}$. By (73), we have $\lambda_v = \lambda/\theta_v \geq \lambda \geq \lambda_2(d_v)$.

In both cases, $\lambda_v \in (0, \lambda_1(d_v)] \cup [\lambda_2(d_v), +\infty)$, and by Proposition 8.6, $(\beta, \gamma, \lambda_v)$ is d_v-unique with gap δ. □

We consider the following well known tree recursions for two-spin systems [?]. Let $0 \leq \beta \leq \gamma, \gamma > 0$ and $\lambda > 0$ be reals (not necessarily satisfying the anti-ferromagnetic requirement $\beta\gamma < 1$). For integer $d \geq 0$ and real $\lambda > 0$, the *tree recursion for marginal-ratios* $F_{\lambda,d} : [0, +\infty]^d \to [0, +\infty]$ is defined as

$$(75) \qquad F_{\lambda,d}(x_1, x_2, \cdots, x_d) \triangleq \lambda \prod_{i=1}^{d} \frac{\beta x_i + 1}{\gamma + x_i}.$$

In particular, $F_{\lambda,0} \triangleq \lambda$ is a constant for the trivial case $d = 0$. The *tree recursion for log-marginal-ratios* $H_{\lambda,d} : [-\infty, +\infty]^d \to [-\infty, +\infty]$ is given by $H_{\lambda,d} = \log \circ F_{\lambda,d} \circ \exp$, specifically

$$(76) \qquad H_{\lambda,d}(y_1, \cdots, y_d) \triangleq \log \lambda + \sum_{i=1}^{d} \log\left(\frac{\beta e^{y_i} + 1}{e^{y_i} + \gamma}\right).$$

For $y \in [-\infty, +\infty]$, let

$$(77) \qquad h(y) \triangleq -\frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}.$$

It holds that $\frac{\partial}{\partial y_i} H_{\lambda,d}(y_1, \cdots, y_d) = h(y_i)$ for all $1 \leq i \leq d$. Note that for $\beta > 0$, $h(y) = 0$ iff $y = \pm\infty$; and for $\beta = 0$, $h(y) = 0$ iff $y = -\infty$.

Given a function $\phi : [-\infty, +\infty] \to [0, +\infty)$ such that $\phi(y) > 0$ for any $y \in [-\infty, +\infty]$ that $|h(y)| > 0$, let $h^\phi : [-\infty, +\infty] \to [0, +\infty)$ be defined as that for any $y \in [-\infty, +\infty]$, $h^\phi(y) = 0$ if $h(y) = 0$, and if $h(y) \neq 0$,

$$(78) \qquad h^\phi(y) = \frac{|h(y)|}{\phi(y)}.$$

Furthermore, let $J_{\lambda,d} \triangleq H_{\lambda,d}[-\infty, +\infty]^d$ denote the image of $H_{\lambda,d}$. Specifically, $J_{\lambda,0} = \{\log \lambda\}$ for $d = 0$ and if $d > 0$,

$$J_{\lambda,d} = \begin{cases} \left[-\infty, \log\left(\frac{\lambda}{\gamma^d}\right)\right] & \text{if } \beta = 0; \\ \left[\log\left(\lambda\beta^d\right), \log\left(\frac{\lambda}{\gamma^d}\right)\right] & \text{if } 0 < \beta\gamma \leq 1; \\ \left[\log\left(\frac{\lambda}{\gamma^d}\right), \log\left(\lambda\beta^d\right)\right] & \text{if } \beta\gamma > 1, \end{cases}$$

Definition 8.7 ((α, c)-potential function). Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system with local fields, where $0 \leq \beta \leq \gamma, \gamma > 0$, and $\lambda_v > 0$ for all $v \in V$. For every $v \in V$, let $d_v \triangleq \Delta_v - 1$, where Δ_v denotes the degree of v in $G = (V, E)$. Let $\phi : [-\infty, +\infty] \to [0, +\infty)$ be a function such that $\phi(y) > 0$ for any $y \in [-\infty, +\infty]$ that $|h(y)| > 0$. For any $\alpha \in (0, 1)$ and $c > 0$, we say ϕ is an (α, c)-potential function with respect to $\mathcal{I}$ if it satisfies:

(1) (α-Contraction) For every $v \in V$ with $d_v \geq 1$ and every $(y_1, \ldots, y_{d_v}) \in [-\infty, +\infty]^{d_v}$, we have

$$\phi(y) \sum_{i=1}^{d_v} h^\phi(y_i) \leq 1 - \alpha.$$

where $y = H_{\lambda_v, d_v}(y_1, \ldots, y_d)$.

(2) (c-Boundedness) For every $u, v \in V$, every $y_u \in J_{\lambda_u, d_u}$ and $y_v \in J_{\lambda_v, d_v}$, we have

$$\phi(y_v) \cdot h^\phi(y_u) \leq \frac{2c}{\Delta_u + \Delta_v}.$$

Theorem 8.8 ([?]). *Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system with local fields, where $0 \leq \beta \leq \gamma$, $\gamma > 0$, and $\lambda_v > 0$ for all $v \in V$. For $\alpha \in (0, 1)$ and $c > 0$, if there is an (α, c)-potential function ϕ with respect to $\mathcal{I}$, then the Gibbs distribution π of $\mathcal{I}$ is $\frac{2c}{\alpha}$-spectrally independent.*

Remark 8.9. The potential function ϕ in Definition 8.7 is in fact the derivative of the potential function in [? , Definition 4]. Theorem 8.8 holds without assuming that ϕ has an explicitly defined integration.

Remark 8.10. Theorem 8.8 holds for general two-spin systems that are not necessarily anti-ferromagnetic. The theorem stated here is in fact a refinement from the one proved in [?] to the two-spin systems with local fields. In Appendix B.2, we reprove the theorem by going through the analyses in [?]. Another slight difference is that our spectral independence is defined with the absolute influence matrix instead of signed influence matrix. This is not an issue, because in the proof of Theorem 8.8, the spectral independence is guaranteed by establishing a sufficient condition of weighted total influences, which is sufficient to imply spectral independence with either absolute or signed matrix.

It remains to verify the contraction and boundedness properties for the instance $\mathcal{I}_\pi = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$, where the local fields $(\lambda_v)_{v \in V}$ are as specified in (74).

For anti-ferromagnetic 2-spin systems, a good potential function ϕ is the one discovered in [?]:

$$(79) \qquad \phi(y) = \sqrt{|h(y)|}.$$

For such choice of potential function it obviously holds that $\phi(y) > 0$ for any $y \in [-\infty, +\infty]$ that $|h(y)| > 0$, and moreover, for any $y \in [-\infty, +\infty]$,

$$h^\phi(y) = \phi(y) = \sqrt{|h(y)|}.$$

Lemma 8.11 ([?]). *Let $d \geq 1$ be an integer, and let β, γ, λ be real numbers satisfying that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\lambda > 0$ and $\beta\gamma < 1$. For the function ϕ defined in (79), for any $\delta \in (0, 1)$, if (β, γ, λ) is d-unique with gap δ, then for every $(y_1, \cdots, y_d) \in [-\infty, +\infty]^d$ and $y = H_{\lambda, d}(y_1, y_2, \cdots, y_d)$,*

$$(80) \qquad \phi(y) \sum_{i=1}^{d} h^\phi(y_i) = \sum_{i=1}^{d} \sqrt{|h(y)| \, |h(y_i)|} \leq \sqrt{1 - \delta} < 1 - \frac{\delta}{2}.$$

Remark 8.12. It was proved in [?] that (80) holds for all integers $1 \leq d < \Delta$ if (β, γ, λ) is up-to-Δ unique, which was in fact proved by showing that (80) holds if (β, γ, λ) is d-unique, which is the exact statement in Lemma 8.11. A proof of Lemma 8.11 is included in Appendix B.3 for completeness,.

Assume that (β, γ, λ) is up-to-Δ unique with gap δ. It follows immediately from Lemma 8.5 and Lemma 8.11 that the potential function ϕ defined in (79) satisfies $\frac{\delta}{2}$-contraction with respect to instance $\mathcal{I}_\pi$.

The following lemma for boundedness was proved in [?].

Lemma 8.13 ([?]). *Let $\Delta \geq 3$ be an integer, and let β, γ, λ be real numbers satisfying that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\lambda > 0$ and $\beta\gamma < 1$. For the potential function ϕ defined in (79), for any $\delta \in (0, 1)$, if (β, γ, λ) is up-to-Δ unique with gap δ, then for any integers $0 \leq d_1, d_2 \leq \Delta - 1$, for every $y_1 \in J_{\lambda, d_1}$ and $y_2 \in J_{\lambda, d_2}$,*

$$\phi(y_1) \cdot h^\phi(y_2) = \sqrt{|h(y_1)| \, |h(y_2)|} \leq \frac{72}{d_1 + d_2 + 2}.$$

Now, to verify the boundedness for ϕ with respect to $\mathcal{I}_\pi$, it is sufficient to show that for every $v \in V$,

(81)
$$\max_{y \in J_{\lambda_v, d_v}} |h(y)| \leq \max_{y \in J_{\lambda, d_v}} |h(y)|.$$

Note that the maximum on the right-hand-side is taken over J_{λ, d_v} and the maximum on the left-hand-side is taken over J_{λ_v, d_v} where $\lambda_v = \theta_v^{\chi_v}$ is the local field associated with $v \in V$ in $\mathcal{I}_\pi$, specifically

$$\lambda_v = \begin{cases} \lambda \cdot \theta_v & \lambda \leq \left(\frac{\gamma}{\beta}\right)^{\Delta_v/2}, \\ \lambda/\theta_v & \text{otherwise}, \end{cases}$$

for an arbitrarily fixed $\theta_v \in (0, 1)$.

When $\beta = 0$, $|h(y)| = \frac{e^y}{e^y + \gamma}$ is monotonically increasing in y, and for every $v \in V$, we have $\lambda_v = \lambda \cdot \theta_v < \lambda$ and hence $\max J_{\lambda_v, d_v} \leq \max J_{\lambda, d_v}$. It holds that

$$\max_{y \in J_{\lambda_v, d_v}} |h(y)| = \left|h(\max J_{\lambda_v, d_v})\right| \leq \left|h(\max J_{\lambda, d_v})\right| = \max_{y \in J_{\lambda, d_v}} |h(y)|.$$

The inequality (81) follows.

In the following we assume $\beta > 0$, i.e. $0 < \beta\gamma < 1$. Note that $|h(y)| = \frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}$ is monotonically increasing in y when $y \leq \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$, and monotonically decreasing in y when $y \geq \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$, so that $|h(y)|$ achieves the maximum at $y = \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$. Then, we can verify (81) by considering three cases:

(1) $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) \in J_{\lambda, d_v}$;
(2) $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) > \log\left(\frac{\lambda}{\gamma^{d_v}}\right)$;
(3) $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) < \log\left(\lambda\beta^{d_v}\right)$.

For Case.1, (81) holds trivially because $\max_{y \in J_{\lambda, d_v}} |h(y)|$ achieves the global maximum of $|h(y)|$.

We then verify (81) in Case.2. Notice that when $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) > \log\left(\frac{\lambda}{\gamma^{d_v}}\right)$, it holds that

$$\lambda < \gamma^{d_v}\left(\frac{\gamma}{\beta}\right)^{\frac{1}{2}} = \left(\frac{\gamma}{\beta}\right)^{\frac{d_v + 1}{2}}(\beta\gamma)^{\frac{d_v}{2}} \leq \left(\frac{\gamma}{\beta}\right)^{\Delta_v/2},$$

which implies $\lambda_v = \lambda \cdot \theta_v < \lambda$ and hence $\max J_{\lambda_v, d_v} \leq \max J_{\lambda, d_v} = \log\left(\frac{\lambda}{\gamma^{d_v}}\right) < \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$. Since $|h(y)|$ is monotonically increasing in y when $y \leq \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$, it holds that

$$\max_{y \in J_{\lambda_v, d_v}} |h(y)| = \left|h(\max J_{\lambda_v, d_v})\right| \leq \left|h(\max J_{\lambda, d_v})\right| = \max_{y \in J_{\lambda, d_v}} |h(y)|.$$

Finally, we verify (81) in Case.3. Notice that when $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) < \log\left(\lambda\beta^{d_v}\right)$, it holds that

$$\lambda > \beta^{-d_v}\left(\frac{\gamma}{\beta}\right)^{\frac{1}{2}} = \left(\frac{\gamma}{\beta}\right)^{\frac{d_v + 1}{2}}(\beta\gamma)^{-\frac{d_v}{2}} \geq \left(\frac{\gamma}{\beta}\right)^{\Delta_v/2},$$

which implies $\lambda_v = \lambda/\theta_v > \lambda$, and hence $\frac{1}{2}\log\left(\frac{\gamma}{\beta}\right) < \log\left(\lambda\beta^{d_v}\right) = \min J_{\lambda, d_v} \leq \min J_{\lambda_v, d_v}$. Since $|h(y)|$ is monotonically decreasing in y when $y \geq \frac{1}{2}\log\left(\frac{\gamma}{\beta}\right)$, it holds that

$$\max_{y \in J_{\lambda_v, d_v}} |h(y)| = \left|h(\min J_{\lambda_v, d_v})\right| \leq \left|h(\min J_{\lambda, d_v})\right| = \max_{y \in J_{\lambda, d_v}} |h(y)|.$$

Therefore, we prove (81). It immediately follows from Lemma 8.13 and (81) that the potential function ϕ defined in (79) satisfies 36-boundedness with respect to instance $\mathcal{I}_\pi$.

Along with the $\frac{\delta}{2}$-contraction we have established for ϕ with respect to $\mathcal{I}_\pi$, this guarantees that ϕ is a $(\frac{\delta}{2}, 36)$-potential function with respect to $\mathcal{I}_\pi$. Due to Theorem 8.8, the distribution π is $\frac{144}{\delta}$-spectrally independent. Note that this holds for $\pi = \mu^{(\theta \chi)}$ for arbitrary $\theta \in (0,1)^V$. This proves the complete $\frac{144}{\delta}$-spectral independence of μ in direction χ claimed in the first half of Lemma 8.4.

8.2. Spectral gap in an easier regime.

We now prove the second half of Lemma 8.4.

Let $\delta \in (0,1)$. Assume that the anti-ferromagnetic 2-spin system $\mathcal{I} = (V, E, \beta, \gamma, \lambda)$ is up-to-Δ unique with gap δ. Fix $\theta = \delta^2/64$ and $C = 8/\delta$. Let

$$\pi = \mu^{(\theta \chi)},$$

where μ is the Gibbs distribution associated with $\mathcal{I}$, and $\chi \in \{-1, +1\}^V$ is the good direction defined in (73).

Our goal is to show that

$$(82) \qquad \lambda_{\text{min-gap}}^{\text{GD}}(\pi) \geq \frac{1}{Cn}.$$

where $\lambda_{\text{min-gap}}^{\text{GD}}(\pi) \triangleq \min_{\Lambda \subseteq V, \sigma_\Lambda \in \Omega(\pi_\Lambda)} \lambda_{\text{gap}}^{\text{GD}}(\pi^{\sigma_\Lambda})$ denotes the minimum spectral gap of the Glauber dynamics for π with worst-case feasible boundary condition.

Let $\Lambda \subseteq V$ be a subset of vertices and $\sigma \in \Omega(\pi_\Lambda)$ a feasible partial configuration specified on Λ. Let P_σ denote the Glauber dynamics for π^σ. We consider the following coupling of chain P_σ.

For any two configurations $X, Y \in \{-1, +1\}^V$, we use $\Phi(X, Y)$ to denote the *weighted hamming distance* between X and Y. Formally

$$\Phi(X, Y) \triangleq \sum_{v \in V : X_v \neq Y_v} \Phi_v,$$

where for each $v \in V$, Φ_v is defined by

$$\Phi_v \triangleq \begin{cases} 1 - \frac{\delta}{8} & , \Delta_v = 1, \\ \Delta_v & , \Delta_v > 1, \end{cases}$$

where Δ_v denotes the degree of v in graph $G = (V, E)$.

Lemma 8.14. *Assume that $\mathcal{I}$ is up-to-Δ unique with gap δ. For any $X, Y \in \Omega(\pi^\sigma)$, there is a coupling $(X, Y) \to (X', Y')$ of Markov chain P_σ such that*

$$\mathbf{E}\left[\Phi(X', Y') \mid X, Y\right] \leq \left(1 - \frac{\delta}{8n}\right) \Phi(X, Y).$$

Due to the well known connection between coupling and spectral gap (Lemma 3.4), our desired spectral gap bound in (82) is implied by Lemma 8.14, and hence this proves the second half of Lemma 8.4.

Remark 8.15. Usually when using coupling to analyze the mixing time upper bound, the metric $\Phi(\cdot, \cdot)$ is required to satisfy $\min_{X, Y \in \Omega} \Phi(X, Y) \geq 1$. Our weighted Hamming distance per se does not satisfy this. This does not matter though, because our purpose of using coupling here is not to bound the mixing time, but the spectral gap, and Lemma 3.4 holds as long as there is sufficient decay in the metric. If one wants to analyze the mixing time using our coupling, one can simply renormalize by replacing Φ with 2Φ, so that the minimum distance is bounded from below and the step-wise decay in Lemma 8.14 still holds.

It only remains to prove Lemma 8.14.

Proof of Lemma 8.14. Observe that $\pi = \mu^{(\theta \chi)}$ corresponds to the Gibbs distribution of a 2-spin system where the pairwise interactions on edges are the same as in μ, but now each $v \in V$ has a local field:

$$\lambda_v \triangleq \lambda \theta^{\chi_v}.$$

More specifically, for every $\tau \in \{-1, +1\}^V$,

$$\pi(\tau) \propto \beta^{m_{+1}(\tau)} \gamma^{m_{-1}(\tau)} \prod_{v:\tau(v)=+1} \lambda_v,$$

where $m_i(\tau) \triangleq |\{(u, v) \in E \mid \tau_u = \tau_v = i\}|$ for $i \in \{-1, +1\}$.

The coupling is constructed by the path coupling [?]. Note that P_σ is a Markov over $\Omega(\pi^\sigma)$. To apply the path coupling, we first extend P_σ to the entire space $\{-1, +1\}^V$. Suppose the current configuration is $X \in \{-1, +1\}^V$, not necessarily feasible with respect to π. Upon transition, a vertex $v \in V$ is picked uniformly at random; if $v \in \Lambda$, then the value of v is fixed by σ and set $X_v \leftarrow \sigma_v$; if $v \in V \setminus \Lambda$, then the current value of X_v is updated to a random value $c_v \in \{-1, +1\}$ such that

(83)
$$\Pr[c_v = -1] = \hat{\mu}_v^{X_{V\setminus\{v\}}}(-1) = p_v(-1, s) \triangleq \frac{\gamma^s}{\gamma^s + \lambda_v \beta^{\Delta_v - s}},$$

$$\Pr[c_v = +1] = \hat{\mu}_v^{X_{V\setminus\{v\}}}(+1) = p_v(+1, s) \triangleq \frac{\lambda_v \beta^{\Delta_v - s}}{\gamma^s + \lambda_v \beta^{\Delta_v - s}},$$

where $s \triangleq |\{u \mid v \in \Gamma_v \wedge X_v = -1\}|$ denotes the number of -1's assigned by X to the neighborhood Γ_v of v in G, and the value of $\beta^{\Delta_v - s}$ is computed with convention $0^0 = 1$.[2]

In (83), $\hat{\mu}$ extends the definition of conditional distribution induced by μ to the boundary conditions that may be infeasible in general. For any feasible configuration X, it is easy to see that $\hat{\mu}_v^{X_{V\setminus\{v\}}} = \mu_v^{X_{V\setminus\{v\}}}$.

Let $X, Y \in \{-1, +1\}^V$ be disagreeing with each other at only one vertex $v \in V$. The coupling $(X, Y) \rightarrow (X', Y')$ is constructed as follows:

- the two chains pick the same vertex $w \in V$ uniformly at random, and $X'_u = Y'_u$ for all $u \neq w$;
- (X'_w, Y'_w) is drawn according to the optimal coupling of their marginal distributions in (83).

If the Glauber dynamics picks v, then $X'_v = Y'_v$; otherwise, $X'_v \neq Y'_v$. We have $\Pr[X'_v \neq Y'_v \mid X, Y] = 1 - \frac{1}{n}$. For any $w \notin \Gamma_v \cup \{v\}$, it holds that $\Pr[X'_w \neq Y'_w \mid X, Y] = 0$. For any $u \in \Gamma_v$, $X'_u \neq Y'_u$ only if the Glauber dynamics picks u and the coupling on vertex u fails. We have

(84)
$$\forall u \in \Gamma(v), \quad \Pr[X'_u \neq Y'_u \mid X, Y] \leq \frac{1}{n} R(u, v),$$

where $R(v, u)$ corresponds to the Dobrushin's influence matrix, formally defined as follows:

$$R(v, u) \triangleq \max_{(\sigma, \tau) \in B_v} d_{\mathrm{TV}}\left(\hat{\mu}_u^{\sigma_{V\setminus\{u\}}}, \hat{\mu}_u^{\tau_{V\setminus\{u\}}}\right),$$

where B_v denotes the set of all pairs $(\sigma, \tau) \in \{-1, +1\}^V \times \{-1, +1\}^V$ that disagree only at v. We have

$$R(v, u) = \max_{0 \leq s \leq \Delta_u - 1} |p_u(+1, s+1) - p_u(+1, s)|$$

$$= \max_{0 \leq s \leq \Delta_u - 1} \frac{\lambda_u \beta^{\Delta_u - s - 1} \gamma^s (1 - \beta\gamma)}{(\gamma^{s+1} + \lambda\beta^{\Delta_u - s - 1})(\gamma^s + \lambda_u \beta^{\Delta_u - s})}$$

$$= \max_{0 \leq s \leq \Delta_u - 1} \frac{\lambda_u \beta^s \gamma^{\Delta_u - s - 1}(1 - \beta\gamma)}{(\gamma^{\Delta_u - s} + \lambda\beta^s)(\gamma^{\Delta_u - s - 1} + \lambda_u \beta^{s+1})} \qquad \text{(by replacing } s \text{ with } \Delta_u - s - 1)$$

$$= \max_{0 \leq s \leq \Delta_u - 1} \frac{\lambda_u \beta^s \gamma^{-\Delta_u + s + 1}(1 - \beta\gamma)}{(\gamma + \lambda\beta^s \gamma^{-\Delta_u + s + 1})(1 + \lambda_u \beta^{s+1} \gamma^{-\Delta_u + s + 1})} \qquad \text{(since } \gamma > 0)$$

$$= \max_{0 \leq s \leq \Delta_u - 1} \frac{1}{\Delta_u} f_{\Delta_u}\left(\frac{\lambda_u (\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right),$$

---

[2]This convention is assumed throughout the proof without further mentioning.

where the function f_d for integer d is as defined in (9):

$$f_d(x) = \frac{d(1 - \beta\gamma)x}{(\beta x + 1)(x + \gamma)}.$$

Altogether, we have

(85)
$$\mathbf{E}\left[\Phi(X', Y') \mid X, Y\right] \leq \Phi_v\left(1 - \frac{1}{n}\right) + \frac{1}{n}\sum_{u \in \Gamma_v} \frac{\Phi_u}{\Delta_u} f_{\Delta_u}\left(\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right).$$

Remark 8.16. Note that $f_{\Delta_u}(\cdot)$ captures the contraction of the tree recursion for the marginal ratio of $(\Delta_u + 1)$-regular tree. We know that $f_{\Delta_u} < 1$ at the fixed point for the tree recursion, if we had assumed the $(\Delta_u + 1)$-uniqueness. However, we only assume the up-to-Δ uniqueness of (β, γ, λ). Such discrepancy is due to the non-self-avoiding nature of path coupling argument. Nevertheless, since we have moved to an easier regime $\pi = \mu^{(\theta^\mathcal{X})}$ that effectively alters the local fields from λ to $\lambda_v = \lambda\theta^{\mathcal{X}_v}$, we could hope for that f_{Δ_u} still contracts at $\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}$ as long as it is fairly close to the fixed point.

The above intuition is formally justified by the following claim.

Claim 8.17. *For any $u \in V$, if $\Delta_u > 1$, then for any integer $0 \leq s \leq \Delta_u - 1$, it holds that*

$$f_{\Delta_u}\left(\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right) < 1 - \frac{\delta}{4}.$$

Claim 8.17 will be proved later. We now use Claim 8.17 to bound (85). Without loss of generality, we assume the underlying graph $G = (V, E)$ is connected. Otherwise, we can decompose the spin system $\mathcal{I}$ into a set of independent systems. Consider the following two cases:

- Case $\Delta_v = 1$. We have $\Phi_v = 1 - \frac{\delta}{8}$. Let $\Gamma_v = \{u\}$. If $\Delta_u = 1$, since G is connected, G only contains two vertices, such instance is trivial. If $\Delta_u > 1$, it holds that $\Phi_u = \Delta_u$. By (85) and Claim 8.17,

$$\mathbf{E}\left[\Phi(X', Y') \mid X, Y\right] \leq \left(1 - \frac{\delta}{8}\right)\left(1 - \frac{1}{n}\right) + \frac{1}{n}\left(1 - \frac{\delta}{4}\right)$$

$$\leq \left(1 - \frac{\delta}{8n}\right)\Phi_v.$$

- Case $\Delta_v > 1$. We have $\Phi_v = \Delta_v$. Fix any neighbor $u \in \Gamma_v$. If $\Delta_u > 1$, then $\Phi_u = \Delta_u$, by Claim 8.17,

$$\frac{\Phi_u}{\Delta_u} f_{\Delta_u}\left(\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right) \leq 1 - \frac{\delta}{4}.$$

If $\Delta_u = 1$, then $\Phi_u = 1 - \frac{\delta}{8}$, by the definition of $R(v, u)$, it is straightforward to verify that

$$R(v, u) = \frac{1}{\Delta_u} f_{\Delta_u}\left(\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right) \leq 1.$$

Thus, we have

$$\frac{\Phi_u}{\Delta_u} f_{\Delta_u}\left(\frac{\lambda_u(\beta\gamma)^s}{\gamma^{\Delta_u - 1}}\right) \leq \Phi_u = 1 - \frac{\delta}{8}.$$

Therefore, by (85), we have

$$\mathbf{E}\left[\Phi(X', Y') \mid X, Y\right] \leq \Delta_v\left(1 - \frac{1}{n}\right) + \frac{\Delta_v}{n}\left(1 - \frac{\delta}{8}\right)$$

$$= \left(1 - \frac{\delta}{8n}\right)\Phi_v.$$

44

Recall that we have extended the Markov chain P_σ to the entire space $\{-1, +1\}^V$. Since the weighted Hamming distance is a well-defined metric on $\{-1, +1\}^V$, due to the path coupling theorem [?], for any $X, Y \in \{-1, +1\}^V$, there is a coupling $(X, Y) \to (X', Y')$ of P_σ such that

$$\mathbf{E}\left[\Phi(X', Y') \mid X, Y\right] \le \left(1 - \frac{\delta}{8n}\right)\Phi_v.$$

This step-wise decay property holds for every $X, Y \in \{-1, +1\}^V$, thus it holds for every $X, Y \in \Omega(\pi^\sigma)$. This proves the lemma. $\qquad\square$

It remains to prove Claim 8.17.

Proof of Claim 8.17. Let $\delta \in (0, 1)$. Fix a vertex $u \in V$ with $\Delta_u \ge 2$. The effective local field of u in π is given by $\lambda_u = \theta^{\chi_u}$, where $\theta = \frac{\delta^2}{64}$, and $\chi_u = \mathrm{sgn}(\Delta_u, \lambda)$ is as defined in (73):

$$\mathrm{sgn}(\Delta_u, \lambda) \triangleq \begin{cases} +1 & \lambda \le \left(\frac{\gamma}{\beta}\right)^{\Delta_u/2} \\ -1 & \text{otherwise.} \end{cases}$$

We denote

$$D = \Delta_u \quad \text{and} \quad d = D - 1.$$

Clearly, $2 \le D \le \Delta$ and $1 \le d \le \Delta - 1$.

Recall that (β, γ, λ) is up-to-Δ unique with gap δ, which means that (β, γ, λ) is d-unique with gap δ. To prove the claim, it suffices to prove the following proposition: For any (β, γ, λ) with $0 \le \beta \le \gamma$, $\beta\gamma < 1$ and $\gamma, \lambda > 0$, for any integer $d \ge 1$, if (β, γ, λ) is d-unique with gap δ, i.e. the following holds:

(86)
$$f_d(x_d) = \frac{d(1 - \beta\gamma)x_d}{(\beta x_d + 1)(x_d + \gamma)} \le 1 - \delta,$$

where $x_d = F(x_d)$ is the unique positive fixed point for $F_d(x) \triangleq \lambda \left(\frac{\beta x + 1}{x + \gamma}\right)^d$, then for $D = d + 1$,

(87)
$$\max_{0 \le s \le D-1} f_D\left(\frac{\lambda \theta^{\mathrm{sgn}(D,\lambda)}(\beta\gamma)^s}{\gamma^d}\right) < 1 - \frac{\delta}{4}.$$

We first consider the case $\beta = 0$. In this case, $f_d(x) = \frac{dx}{x + \gamma}$. Hence, $f_d(x) < 1 - \delta$ if and only if $x < \frac{(1-\delta)\gamma}{d-1+\delta}$. And the fixed point $x_d = F(x_d)$ satisfies $\lambda = x_d(x_d + \gamma)^d$. Thus, the assumption (86) implies that

(88)
$$\lambda = x_d(x_d + \gamma)^d \le \left(\frac{(1-\delta)\gamma}{d-1+\delta}\right)\left(\frac{(1-\delta)\gamma}{d-1+\delta} + \gamma\right)^d = \frac{(1-\delta)\gamma^{d+1}d^d}{(d-1+\delta)^{d+1}}.$$

Note that in this case, $\mathrm{sgn}(D, \lambda) = 1$ for all integer $D \ge 2$ and real number $\lambda > 0$. Also note that if $s \ge 1$, then $(\beta\gamma)^s = 0$ and $f_D(0) = 0$. We then only need to prove Claim 8.17 when $s = 0$. Formally, this is equivalent to show that

$$f_D\left(\frac{\theta\lambda}{\gamma^d}\right) < 1 - \frac{\delta}{4}.$$

Note that $f_D(x) = \frac{Dx}{x + \gamma}$ when $\beta = 0$. Hence, $f_D(x) < 1 - \frac{\delta}{4}$ if and only if $x < \frac{\gamma(4-\delta)}{4(D-1)+\delta} = \frac{\gamma(4-\delta)}{4d+\delta}$. We then only need to verify

$$\theta\lambda < \frac{\gamma^{d+1}(1 - \frac{\delta}{4})}{d + \frac{\delta}{4}}.$$

45

Note that $d \geq 1$ and $0 < \delta < 1$. By (88), it suffices to verify

$$\theta = \frac{\delta^2}{64} \leq \frac{1}{2} \cdot \left(\frac{d-1+\delta}{d}\right)^{d+1}.$$

Let $h(t) = \left(\frac{t-1+\delta}{t}\right)^{t+1}$. The above inequality can be expressed as $h(d) \geq \frac{\delta^2}{32}$, which holds because $h(t)$ is increasing in t for $t \geq 1$ and $h(1) = \delta^2$. To verify the increasing of $h(t)$, observe that

$$(\ln h(t))' = \ln\left(1 - \frac{1-\delta}{t}\right) + \frac{(1-\delta)(t+1)}{t(t-1+\delta)} \to 0 \text{ as } t \to \infty,$$

$$(\ln h(t))'' = \frac{(1-\delta)(-3t+\delta t+1-\delta)}{t^2(t-1+\delta)^2} < 0 \text{ if } t \geq 1.$$

Next, we focus on the main case $\beta > 0$. We first show that without loss of generality, we can assume $\lambda \leq (\gamma/\beta)^{D/2}$. Suppose $\lambda > (\gamma/\beta)^{D/2}$. Let $\lambda' = \frac{1}{\lambda} \cdot (\gamma/\beta)^D$. The following two properties hold:

- if (β, γ, λ) is d-unique with gap δ, then $(\lambda', \beta, \gamma)$ is also d-unique with gap δ;
- $\lambda' < (\gamma/\beta)^{D/2}$ and it holds that

(89)
$$\max_{0 \leq s < D} f_D\left(\frac{\lambda \theta^{\mathrm{sgn}(D,\lambda)}(\beta\gamma)^s}{\gamma^d}\right) = \max_{0 \leq s < D} f_D\left(\frac{\lambda' \theta^{\mathrm{sgn}(D,\lambda')}(\beta\gamma)^s}{\gamma^d}\right).$$

We verify the first property. If $d < (1-\delta)\overline{\Delta}$, then by Proposition 8.6, the property holds trivially. Now suppose $d \geq (1-\delta)\overline{\Delta}$. By Proposition 8.6, $\lambda_1(d)\lambda_2(d) = (\gamma/\beta)^D$. Since (β, γ, λ) is d-unique with gap δ and $\lambda > (\gamma/\beta)^{D/2}$, we have $\lambda > \lambda_2(d)$ and $\lambda' = \frac{1}{\lambda} \cdot (\gamma/\beta)^D < \lambda_1(d)$, thus $(\lambda', \beta, \gamma)$ is also d-unique with gap δ.

We then verify the second property. Since $\lambda > (\gamma/\beta)^{D/2}$, $\lambda' = \frac{1}{\lambda} \cdot (\gamma/\beta)^D < (\gamma/\beta)^{D/2}$. To verify (89), observe that the following holds for the function f_D:

$$\forall x > 0, \quad f_D(x) = \frac{D(1-\beta\gamma)x}{(\beta x + 1)(x+\gamma)} = f_D\left(\frac{\gamma}{x\beta}\right).$$

This implies

$$\max_{0 \leq s < D} f_D\left(\frac{\lambda \theta^{\mathrm{sgn}(D,\lambda)}(\beta\gamma)^s}{\gamma^d}\right) = \max_{0 \leq s < D} f_D\left(\frac{\gamma}{\beta} \cdot \frac{\gamma^d}{\lambda \theta^{\mathrm{sgn}(D,\lambda)}(\beta\gamma)^s}\right)$$

$$= \max_{0 \leq s < D} f_D\left(\frac{\lambda' \theta^{\mathrm{sgn}(D,\lambda')}(\beta\gamma)^{d-s}}{\gamma^d}\right)$$

$$= \max_{0 \leq s < D} f_D\left(\frac{\lambda' \theta^{\mathrm{sgn}(D,\lambda')}(\beta\gamma)^s}{\gamma^d}\right).$$

With the above two properties, we only need to prove the following result. Let $\delta \in (0, 1)$. For any (β, γ, λ), any integer $d \geq 1$ and $D = d+1$, if (β, γ, λ) is d-unique with gap δ (formally, (86)) and $\lambda \leq (\gamma/\beta)^{D/2}$, then it holds that

(90)
$$\forall \text{ integer } 0 \leq s \leq D - 1, \quad f_D\left(\frac{\lambda \theta^{\mathrm{sgn}(D,\lambda)}(\beta\gamma)^s}{\gamma^d}\right) = f_D\left(\frac{\lambda \theta(\beta\gamma)^s}{\gamma^d}\right) \leq 1 - \frac{\delta}{4}.$$

Furthermore, we have the following technical proposition.

Proposition 8.18. *For integer $d \geq (1-\delta)\overline{\Delta}$, if (β, γ, λ) is d-unique with gap δ, then*

$$\frac{\lambda(\beta\gamma)^s \delta}{\gamma^d(2-\delta)} \leq x_1(d),$$

for any integer $0 \leq s \leq d$, where $\zeta_\delta(d) = d(1-\beta\gamma) - (1-\delta)(1+\beta\gamma)$.

The proposition is proved later.

We now prove (87). Consider two cases: $D \geq \left(1 - \frac{\delta}{2}\right)\overline{\Delta} + 1$ and $D < \left(1 - \frac{\delta}{2}\right)\overline{\Delta} + 1$.

Case 1 ($D \geq (1 - \delta/2)\overline{\Delta} + 1$). In this case, $d = D - 1 \geq (1 - \frac{\delta}{2})\overline{\Delta} \geq (1 - \delta)\overline{\Delta}$, we claim

$$(91) \qquad \beta x_1(d) + \frac{\gamma}{x_1(d)} \geq \frac{\delta}{2}(1 + \beta\gamma),$$

where $x_1(d)$ is defined in Proposition 8.6. Since $\theta = \frac{\delta^2}{64} \leq \frac{\delta}{8}\frac{\delta}{2-\delta}$, we have

$$\frac{\lambda\theta(\beta\gamma)^s}{\gamma^d} \leq \frac{\lambda(\beta\gamma)^s}{\gamma^d} \cdot \frac{\delta}{2-\delta} \cdot \frac{\delta}{8} \overset{(\star)}{\leq} \frac{\delta}{8}x_1(d) \overset{(*)}{\leq} \sqrt{\gamma/\beta}.$$

where $(\star)$ is due to Proposition 8.18 and $(*)$ holds because $x_1(d)x_2(d) = \gamma/\beta$ and $x_1(d) \leq x_2(d)$, thus $x_1(d) \leq \sqrt{\gamma/\beta}$. Furthermore, it is easy to verify that $f_D(x)$ is increasing on $(0, \sqrt{\gamma/\beta}]$. We have

$$f_D\left(\frac{\lambda\theta(\beta\gamma)^s}{\gamma^d}\right) \leq f_D\left(\frac{\delta}{8}x_1(d)\right)$$

$$= \frac{D(1 - \beta\gamma)}{\frac{\delta}{8}\beta x_1(d) + \frac{8}{\delta}\frac{\gamma}{x_1(d)} + 1 + \beta\gamma}$$

$$\leq \frac{D(1 - \beta\gamma)}{\frac{8}{\delta}\frac{\gamma}{x_1(d)} + 1 + \beta\gamma}$$

$$\leq \frac{D(1 - \beta\gamma)}{\frac{4}{\delta}\left(\beta x_1(d) + \frac{\gamma}{x_1(d)}\right) + 1 + \beta\gamma} \qquad \text{(since } x_1(d) \leq \sqrt{\gamma/\beta}\text{)}$$

$$\leq \frac{D(1 - \beta\gamma)}{\frac{2}{\delta}\left(\beta x_1(d) + \frac{\gamma}{x_1(d)}\right) + 1 + \beta\gamma + 2\left(\beta x_1(d) + \frac{\gamma}{x_1(d)}\right)} \qquad \text{(since } 0 < \delta < 1\text{)}$$

$$\leq \frac{D(1 - \beta\gamma)}{2(\beta x_1(d) + \frac{\gamma}{x_1(d)} + 1 + \beta\gamma)} \qquad \text{(by (91))}$$

$$= \frac{D}{2d}f_d(x_1(d))$$

$$\leq (1 - \delta),$$

where the last inequality holds because $x_1(d)$ is the root of $f_d(x) = 1 - \delta$ (due to Proposition 8.6).

We finish the analysis of this case by verifying (91). Note that it holds that

$$(92) \qquad \frac{1 + \beta\gamma}{d(1 - \beta\gamma)} \leq \frac{(1 + \sqrt{\beta\gamma})^2}{d(1 + \sqrt{\beta\gamma})(1 - \sqrt{\beta\gamma})} = \frac{\overline{\Delta}}{d} \overset{(*)}{\leq} \frac{1}{1 - \frac{\delta}{2}},$$

where $(*)$ holds because $D \geq \left(1 - \frac{\delta}{2}\right)\overline{\Delta} + 1$. Moreover, we have

$$(93) \qquad \frac{\beta x_1(d) + \frac{\gamma}{x_1(d)}}{d(1 - \beta\gamma)} + \frac{1 + \beta\gamma}{d(1 - \beta\gamma)} = \frac{1}{f_d(x_1(d))} = \frac{1}{1 - \delta}.$$

Combining (92) and (93), we have

$$\frac{\beta x_1(d) + \frac{\gamma}{x_1(d)}}{1 + \beta\gamma} + 1 \geq \left(\frac{1}{1 - \delta}\right)\left(1 - \frac{\delta}{2}\right) = \frac{\delta}{2 - 2\delta} + 1 \geq \frac{\delta}{2} + 1.$$

Case 2 ($D < (1 - \delta/2)\overline{\Delta} + 1$). Without loss of generality, we assume $\overline{\Delta} \geq \frac{2}{2-\delta}$, since otherwise we have $D \leq \left(1 - \frac{\delta}{2}\right)\overline{\Delta} + 1 < 2$, hence the only possible value for D is 1. However, in (87), we assume $D = d + 1 \geq 2$.

We claim that the following equations hold in this case.

(94)
$$\forall c \in \left(0, \frac{1}{2}\right), \quad f_D\left(c\sqrt{\frac{\gamma}{\beta}}\right) \le \frac{4c \cdot D}{\overline{\Delta} - 1},$$

(95)
$$\frac{\lambda(\beta\gamma)^s}{\gamma^d} \le \sqrt{\frac{\gamma}{\beta}} \cdot \frac{4}{\delta}.$$

Note that $\theta = \frac{\delta^2}{64} = \frac{\delta}{16} \cdot \frac{\delta}{4}$. Then by (95), it holds that
$$\frac{\lambda\theta(\beta\gamma)^s}{\gamma^d} \le \frac{\delta}{16}\sqrt{\frac{\gamma}{\beta}} \le \sqrt{\frac{\gamma}{\beta}}.$$

It is easy to verify that $f_D(x)$ is monotonically increasing on $(0, \sqrt{\gamma/\beta}]$. It holds that
$$f_D\left(\frac{\lambda\theta(\beta\gamma)^s}{\gamma^d}\right) \le f_D\left(\frac{\delta}{16}\sqrt{\frac{\gamma}{\beta}}\right) \overset{(*)}{\le} \frac{\delta D}{4(\overline{\Delta} - 1)} \overset{(\star)}{\le} \frac{\delta}{4} \cdot \frac{\overline{\Delta} + 1}{\overline{\Delta} - 1} \le 1 - \frac{\delta}{4},$$

where $(*)$ follows from (94), $(\star)$ is due to that $D \le \overline{\Delta} + 1$, and the last inequality holds since $\overline{\Delta} \ge \frac{2}{2-\delta}$.

Now, we verify (94), it holds that
$$f_D\left(c\sqrt{\frac{\gamma}{\beta}}\right) \le \frac{D(1 - \beta\gamma)}{\frac{1}{c}\sqrt{\beta\gamma} + 1 + \beta\gamma} = \frac{D(1 - \sqrt{\beta\gamma})}{\frac{(\frac{1}{c}-2)}{1+\sqrt{\beta\gamma}}\sqrt{\beta\gamma} + (1 + \sqrt{\beta\gamma})} \le \frac{D(1 - \sqrt{\beta\gamma})}{\frac{(\frac{1}{c}-2)}{2}\sqrt{\beta\gamma} + (1 + \sqrt{\beta\gamma})},$$

where the last inequality holds because $\sqrt{\beta\gamma} < 1$ and $c < \frac{1}{2}$, which implies
$$f_D\left(c\sqrt{\frac{\gamma}{\beta}}\right) \le \frac{D(1 - \sqrt{\beta\gamma})}{1 + \frac{1}{2c}\sqrt{\beta\gamma}} = \frac{2D}{\left(1 + \frac{1}{2c}\right)\overline{\Delta} + \left(1 - \frac{1}{2c}\right)} \le \frac{2D}{\left(1 + \frac{1}{2c}\right)(\overline{\Delta} - 1)} \le \frac{4c \cdot D}{\overline{\Delta} - 1}.$$

Now we verify (95). It holds that
$$\frac{\lambda(\beta\gamma)^s}{\gamma^d} \le \frac{\lambda}{\gamma^d} \overset{(\star)}{\le} \left(\frac{\gamma}{\beta}\right)^{D/2} \frac{1}{\gamma^d} = \left(\frac{\gamma}{\beta}\right)^{d/2+1/2} \frac{1}{\gamma^d} = \sqrt{\frac{\gamma}{\beta}}\left(\frac{1}{\sqrt{\beta\gamma}}\right)^d,$$

where $(\star)$ is due to our assumption $\lambda \le (\gamma/\beta)^{D/2}$. For $\overline{\Delta} \triangleq \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$, it holds that $\sqrt{\beta\gamma} = \frac{\overline{\Delta}-1}{\overline{\Delta}+1}$, which implies

$$\frac{\lambda(\beta\gamma)^s}{\gamma^d} \le \sqrt{\frac{\gamma}{\beta}}\left(\frac{1}{\sqrt{\beta\gamma}}\right)^d = \sqrt{\frac{\gamma}{\beta}}\left(\frac{\overline{\Delta}+1}{\overline{\Delta}-1}\right)^d \overset{(*)}{\le} \sqrt{\frac{\gamma}{\beta}}\left(\frac{\overline{\Delta}+1}{\overline{\Delta}-1}\right)^{\overline{\Delta}\cdot\frac{2-\delta}{2}} \overset{(\star)}{\le} \sqrt{\frac{\gamma}{\beta}} \cdot \frac{4-\delta}{\delta} \le \sqrt{\frac{\gamma}{\beta}} \cdot \frac{4}{\delta},$$

where $(*)$ is due to that $d \le \left(1 - \frac{\delta}{2}\right)\overline{\Delta}$, and $(\star)$ is due to that the function $h(t) = \left(\frac{t+1}{t-1}\right)^t$ is monotonically decreasing when $t > 1$ and $\overline{\Delta} \ge \frac{2}{2-\delta}$. The monotonicity of function $h(t)$ can be verified as
$$(\ln h(t))' = -\frac{2t}{t^2 - 1} + \ln\left(\frac{t+1}{t-1}\right) \to 0 \text{ as } t \to \infty,$$
$$(\ln h(t))'' = \frac{4}{t^2 - 1} > 0.$$

Combining the two cases proves (87). □

Proof of Proposition 8.18. Recall that $\zeta_\delta(d) \triangleq d(1 - \beta\gamma) - (1 - \delta)(1 + \beta\gamma)$. By Proposition 8.6,
$$x_1(d) = \frac{2(1 - \delta)\gamma}{\zeta_\delta(d) + \sqrt{\zeta_\delta(d)^2 - 4(1 - \delta)^2\beta\gamma}} \le \frac{2(1 - \delta)\gamma}{\zeta_\delta(d)}.$$

Since $\frac{x+\gamma}{\beta x+1}$ is increasing in x when $\beta\gamma < 1$, it holds that

$$\frac{x_1(d) + \gamma}{\beta x_1(d) + 1} \leq \frac{\frac{2(1-\delta)\gamma}{\zeta_\delta(d)} + \gamma}{\frac{2(1-\delta)\beta\gamma}{\zeta_\delta(d)} + 1} = \gamma \cdot \frac{d + (1-\delta)}{d - (1-\delta)},$$

which, by Proposition 8.6, implies

$$\lambda_1(d) = x_1(d)\left(\frac{x_1(d) + \gamma}{\beta x_1(d) + 1}\right)^d \leq x_1(d) \cdot \gamma^d \cdot \left(\frac{d + (1-\delta)}{d - (1-\delta)}\right)^d \overset{(\star)}{\leq} x_1(d) \cdot \gamma^d \cdot \frac{2 - \delta}{\delta},$$

where $(\star)$ holds because $f(t) = \left(\frac{t+(1-\delta)}{t-(1-\delta)}\right)^t$ is decreasing in t when $t \geq 1$ and $d \geq 1$. And thus,

$$\lambda_1(d)\frac{(\beta\gamma)^s}{\gamma^d} \leq \frac{\lambda_1(d)}{\gamma^d} \leq x_1(d) \cdot \frac{2 - \delta}{\delta}.$$

Finally, we note that $f(t)$ is monotonically decreasing because

$$(\ln f(t))' = -\frac{2t(1-\delta)}{(t + (1-\delta))(t - (1-\delta))} + \log\left(\frac{t + (1-\delta)}{t - (1-\delta)}\right) \to 0 \text{ as } t \to +\infty,$$

$$(\ln f(t))'' = \frac{4(1-\delta)^3}{(t + (1-\delta))^2(t - (1-\delta))^2} > 0. \qquad \square$$

8.3. **Hardcore and Ising models.** We now prove Theorem 1.3 and Theorem 1.4, respectively for the hardcore and Ising models. For these models, the constant $C(\delta)$ can be improved to $\exp(O(1/\delta))$ as the parameter θ can be chosen to be a universal constant.

Proof of Theorem 1.3. Without loss of generality, we assume that $\lambda \geq \frac{1}{2\Delta}$, and for $\lambda < \frac{1}{2\Delta}$, the standard path coupling technique gives us $O(n \log n)$ mixing time upper bound and $\left(\frac{1}{2n}\right)$ spectral gap lower bound.
 Choose $\theta = \frac{1}{25}$. We have

$$\theta\lambda = \frac{\lambda}{25} < \frac{(\Delta - 1)^{\Delta-1}}{25(\Delta - 2)^\Delta} < \frac{1}{2\Delta}.$$

Again, by the standard path coupling method, for the hardcore model with fugacity $\theta\lambda < \frac{1}{2\Delta}$, the Glauber dynamics has a $\frac{1}{2n}$ spectral gap lower bound, which further holds up to an arbitrary feasible boundary condition $\sigma_\Lambda \in \Omega(\mu_\Lambda)$ where $\Lambda \subseteq V$. Therefore

$$\lambda_{\text{min-gap}}^{\text{GD}}\left(\mu^{(1/25)}\right) \leq \frac{1}{2n}.$$

On the other hand, by Lemma 8.4, the Gibbs distribution μ is completely $\frac{144}{\delta}$-spectrally independence. It follows from Theorem 1.9 that

$$\lambda_{\text{gap}}^{\text{GD}}(\mu) \geq \left(\frac{1}{50}\right)^{288/\delta+7} \frac{1}{2n} = \frac{1}{C(\delta)n},$$

for a $C(\delta) = \exp(O(1/\delta))$. For $\lambda \geq \frac{1}{2\Delta}$, the marginal bound $b \geq \min\left\{\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\right\} \geq \frac{1}{2\Delta+1}$, which implies that $\mu_{\min} \geq b^n \geq \frac{1}{(2\Delta+1)^n}$. The mixing time bound follows from (1). $\qquad \square$

For the Ising model with edge activity $\beta > 0$, without loss of generality, we can assume $\lambda \leq 1$, because by symmetry, this covers the $\lambda > 1$ case by switching the roles of -1 and $+1$ for all vertices.

Proof of Theorem 1.4. Let μ be the Gibbs distribution of the Ising model on graph $G = (V, E)$ with edge activity $\beta > 0$ and $\lambda \leq 1$. Assume that $\beta \in \left[\frac{\Delta-2+\delta}{\Delta-\delta}, \frac{\Delta-\delta}{\Delta-2+\delta}\right]$, where $\Delta = \Delta_G$ is the maximum degree of G.

We first verify that μ is completely $\frac{4}{\delta}$-spectrally independent. Consider the Ising model with local fields $\mathcal{I}' = (V, E, \beta, (\lambda_v)_{v \in V})$ defined on the same graph $G = (V, E)$ with the same edge activity $\beta \in \left[\frac{\Delta - 2 + \delta}{\Delta - \delta}, \frac{\Delta - \delta}{\Delta - 2 + \delta} \right]$, where each vertex $v \in V$ is associated with an arbitrary local field $\lambda_v \le 1$.

It is straightforward to verify that the trivial potential function $\phi(y) = 1$ is a $(\delta, 2)$-potential function with respect to $\mathcal{I}'$ (Definition 8.7): for $h(y) = -\frac{(1 - \beta^2) e^y}{(\beta e^y + 1)(e^y + \beta)}$,

(1) (δ-contraction) for any $1 \le d \le \Delta - 1$ and $y_1, \ldots, y_d \in [-\infty, +\infty]$,

$$\sum_{i=1}^{d} |h(y_i)| = \sum_{i=1}^{d} \frac{\left| 1 - \beta^2 \right| e^{y_i}}{(\beta e^{y_i} + 1)(e^{y_i} + \beta)} \le \frac{(\Delta - 1) \left| 1 - \beta \right|}{1 + \beta} \le 1 - \delta;$$

(2) (2-boundedness) similarly, for all $y \in [-\infty, +\infty]$, $|h(y)| = \frac{(1 - \beta^2) e^y}{(\beta e^y + 1)(e^y + \beta)} \le \frac{|1 - \beta|}{1 + \beta} \le \frac{1}{\Delta - 1} \le \frac{2}{\Delta}$.

By Theorem 8.8, the Gibbs distribution of such $\mathcal{I}'$ is always $\frac{4}{\delta}$-spectrally independent, which means that μ is completely $\frac{4}{\delta}$-spectrally independent.

On the other hand, the standard path coupling method gives an $O(n \log n)$ mixing time upper bound and a $\frac{1}{2n}$ spectral gap lower bound for $\lambda \le \frac{1}{500}$, which holds up to an arbitrary feasible boundary condition $\sigma_\Lambda \in \Omega(\mu_\Lambda)$ where $\Lambda \subseteq V$. Choose $\theta = \frac{1}{500}$. We have

$$\lambda_{\text{min-gap}}^{\text{GD}} \left(\mu^{(1/500)} \right) \le \frac{1}{2n}.$$

It follows from Theorem 1.9 that

$$\lambda_{\text{gap}}^{\text{GD}}(\mu) \ge \left(\frac{1}{1000} \right)^{8/\delta + 7} \frac{1}{2n} = \frac{1}{C(\delta) n},$$

for a $C(\delta) = \exp(O(1/\delta))$. For $\beta \in \left[\frac{\Delta - 2 + \delta}{\Delta - \delta}, \frac{\Delta - \delta}{\Delta - 2 + \delta} \right]$ and $\frac{1}{500} \le \lambda \le 1$, the marginal bound $b \ge \frac{1}{14000}$ [?], which implies that $\mu_{\min} = b^n \ge \left(\frac{1}{14000} \right)^n$. The mixing time bound follows from (1). $\qquad \square$

Appendix A. Mixing of Uniform Block Dynamics from Spectral Independence

Here we give a short proof of Theorem 2.5 using theorems and lemmas proved in [?, ?]. Let V be a ground set of size n. For $\eta > 0$, μ is an η-spectrally independent distribution over $\{-1, +1\}^V$. For $2\lceil \eta \rceil \leq \ell \leq n$, we want to show that the spectral gap of the uniform ℓ-block dynamics for μ satisfies

$$\lambda_{\text{gap}}(P_\ell) \geq \left(\frac{\ell}{2n} \right)^{2\lceil \eta \rceil + 1}.$$

For any subset $\Lambda \subset V$, any feasible partial configuration $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, define the *local random walk* P_{σ_Λ} on $U_{\sigma_\Lambda} = \left\{ (u, c) \in (V \setminus \Lambda) \times \{-1, +1\} \mid \mu_u^{\sigma_\Lambda}(c) > 0 \right\}$ as

$$\forall (u, i), (v, j) \in U_{\sigma_\Lambda}, \quad P_{\sigma_\Lambda}((u, i), (v, j)) \triangleq \frac{\mathbf{1}[u \neq v]}{|V| - |\Lambda| - 1} \mu_v^{\sigma_\Lambda, u \leftarrow i}(j),$$

where $\mu_v^{\sigma_\Lambda, u \leftarrow i}(j)$ denotes the marginal distribution at v induced from μ conditioned on the configuration on Λ being fixed as σ_Λ and the value of u being fixed as i.

Then, for $\eta > 0$, we say μ is a $(\zeta_0, \cdots, \zeta_{n-2})$-*local spectral expander* if for every $0 \leq k \leq n - 2$, every $\Lambda \subseteq V$ such that $|\Lambda| = k$, and every feasible partial configuration $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, we have

$$\lambda_2(P_{\sigma_\Lambda}) \leq \zeta_k.$$

Theorem A.1 ([?, Theorem A.9]). *Let $\zeta_0, \zeta_1, \ldots, \zeta_{n-2} \in [0, 1]$ be a sequence of positive real numbers. If μ is a $(\zeta_0, \cdots, \zeta_{n-2})$-local spectral expander, then for any $1 \leq \ell \leq n$, it holds that*

$$\lambda_{\text{gap}}(P_\ell) \geq \frac{\sum_{k=n-\ell}^{n-1} \Gamma_k}{\sum_{k=0}^{n-1} \Gamma_k},$$

where $\Gamma_k \triangleq \prod_{j=0}^{k-1} \frac{1 - \zeta_j}{1 + \zeta_j}$ for $k > 0$ and $\Gamma_0 \triangleq 1$.

The following lemma from [?] assumes our notion of spectral independence defined using the absolute influence matrices $\Psi_\mu^{\sigma_\Lambda}$ (Definition 1.5).

Lemma A.2 ([?, Lemma 3.6]). *For $\eta > 0$, if μ is η-spectrally independent, then there is a sequence $\zeta_0, \zeta_1, \cdots, \zeta_{n-2} \in [0, 1]$, such that μ is a $(\zeta_0, \cdots, \zeta_{n-2})$-local spectral expander, where for $0 \leq k \leq n - 2$,*

$$(96) \qquad \zeta_k = \min \left\{ 1, \frac{\eta}{n - k - 1} \right\}.$$

Proof of Theorem 2.5. Let $C \triangleq \lceil \eta \rceil$. Note that μ is C-spectrally independent assuming that μ is η-spectrally independent. By Lemma A.2, for the ζ_k's given in (96), μ is a $(\zeta_0, \cdots, \zeta_{n-2})$-local spectral expander. By some calculation, one can verify that for any $1 \leq k \leq n - 1$, it always holds that

$$\Gamma_k = \prod_{j=0}^{k-1} \frac{1 - \zeta_j}{1 + \zeta_j} = \frac{(n - k - C)(n - k - C + 1) \ldots (n - k + C - 1)}{(n - C)(n + 1 - C) \cdots (n - 1 + C)}.$$

Therefore, by Theorem A.1,

$$\lambda_{\text{gap}}(P_\ell) \geq \frac{\sum_{k=n-\ell}^{n-1} \Gamma_k}{\sum_{k=0}^{n-1} \Gamma_k} \overset{(\star)}{=} \prod_{k=-C}^{C} \frac{\ell + k}{n + k} \geq \left(\frac{\ell - C}{n - C} \right)^{2C+1} \geq \left(\frac{\ell}{2n} \right)^{2C+1},$$

where we use the fact that $n \geq \ell \geq 2C$, and $(\star)$ is due to the equation $\sum_{j=0}^{N-1} \binom{j}{k} = \binom{N}{k+1}$. $\qquad \square$

In this section, we give missing proofs in Section 8.1. Specifically, they are the proofs of Proposition 8.6 (in Appendix B.1), Theorem 8.8 (in Appendix B.2), and Lemma 8.11 (in Appendix B.3). All these theorems have been proved in some variant forms previously. But here our goal is to reestablish them for the two-spin systems with the local fields that are biased towards the easier directions as indicated in (6). In principle, such alternatton should only make these theorems more satisfied. However, to formally verify this, we have to reiterate their respective existing proofs.

B.1. Proof of Proposition 8.6. When $\beta = 0$, recall that (β, γ, λ) is d-unique if and only if

$$(97) \qquad f_d(\hat{x}_d) = \frac{d(1 - \beta\gamma)\hat{x}_d}{(\beta\hat{x}_d + 1)(\gamma + \hat{x}_d)} = \frac{d\hat{x}_d}{\gamma + \hat{x}_d} \leq 1 - \delta,$$

where $\hat{x}_d$ is the fixed point of function $F_d(x) = \lambda\left(\frac{\beta x + 1}{\gamma + x}\right)^d = \lambda\left(\frac{1}{\gamma + x}\right)^d$, namely $F_d(\hat{x}_d) = \hat{x}_d$.

Note that (97) holds if and only if $\hat{x}_d \leq \frac{(1-\delta)\gamma}{d-1+\delta}$, which is equivalent to

$$\lambda \leq \hat{x}_d\left(\frac{\hat{x}_d + \gamma}{\beta\hat{x}_d + 1}\right)^d = \hat{x}_d\left(\hat{x}_d + \gamma\right)^d \leq \frac{(1 - \delta)d^d\gamma^{d+1}}{(d - 1 + \delta)^{d+1}}.$$

We assume that $\beta > 0$. Similar to previous discussion, (β, γ, λ) is d-unique if and only if

$$(98) \qquad f_d(\hat{x}_d) = \frac{d(1 - \beta\gamma)\hat{x}_d}{(\beta\hat{x}_d + 1)(\gamma + \hat{x}_d)} \leq 1 - \delta,$$

where $\hat{x}_d$ is the fixed point of function $F_d(x) = \lambda\left(\frac{\beta x + 1}{\gamma + x}\right)^d$, namely $F_d(\hat{x}_d) = \hat{x}_d$.

When $d < (1 - \delta)\overline{\Delta}$, the following holds for any $x > 0$.

$$f_d(x) = \frac{d(1 - \beta\gamma)x}{(\beta x + 1)(x + \gamma)} = \frac{d(1 - \beta\gamma)}{\beta x + \frac{\gamma}{x} + 1 + \beta\gamma} \leq \frac{d(1 - \beta\gamma)}{1 + \beta\gamma + 2\sqrt{\beta\gamma}} = \frac{d(1 - \sqrt{\beta\gamma})}{1 + \sqrt{\beta\gamma}} < 1 - \delta.$$

Therefore, (β, γ, λ) is d-unique with gap δ if $d < (1 - \delta)\overline{\Delta}$.

If $d \geq (1 - \delta)\overline{\Delta}$, it can be verified that equation $\frac{d(1-\beta\gamma)x}{(\beta x+1)(x+\gamma)} = 1 - \delta$ has two positive roots

$$x_1(d) = \frac{\zeta_\delta(d) - \sqrt{\zeta_\delta(d)^2 - 4(1 - \delta)^2\beta\gamma}}{2(1 - \delta)\beta} \quad \text{and} \quad x_2(d) = \frac{\zeta_\delta(d) + \sqrt{\zeta_\delta(d)^2 - 4(1 - \delta)^2\beta\gamma}}{2(1 - \delta)\beta},$$

and (98) holds if and only if $\hat{x}_d \leq x_1(d)$ or $\hat{x}_d \geq x_2(d)$. Note that $x\left(\frac{x+\gamma}{\beta x+1}\right)^d$ is monotone increasing in x for any fixed d. So, (β, γ, λ) is d-unique with gap δ if and only if

$$\lambda \in (0, \lambda_1(d)] \cup [\lambda_2(d), +\infty),$$

where $\lambda_i(d) = x_i(d)\left(\frac{x_i(d)+\gamma}{\beta x_i(d)+1}\right)^d$, $i \in \{1, 2\}$.

Lastly, $x_1(d)x_2(d) = \frac{\gamma}{\beta}$ by Vieta's formula. Therefore,

$$\begin{aligned}
\lambda_1(d)\lambda_2(d) &= x_1(d)x_2(d)\left(\frac{(\gamma + x_1(d))(\gamma + x_2(d))}{(\beta x_1(d) + 1)(\beta x_2(d) + 1)}\right)^d \\
&= x_1(d)x_2(d)\left(\frac{\gamma^2 + \gamma(x_1(d) + x_2(d)) + x_1(d)x_2(d)}{\beta^2 x_1(d)x_2(d) + \beta(x_1(d) + x_2(d)) + 1}\right)^d \\
&= \frac{\gamma}{\beta}\left(\frac{\gamma^2 + \frac{\gamma}{\beta} + \gamma(x_1(d) + x_2(d))}{\gamma\beta + 1 + \beta(x_1(d) + x_2(d))}\right)^d = \left(\frac{\gamma}{\beta}\right)^{d+1}.
\end{aligned}$$

B.2. Proof of Theorem 8.8. Assuming the existence of (α, c)-potential function ϕ with respect to a two-spin system $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$, we prove that its Gibbs distribution π is $\frac{2c}{\alpha}$-spectrally independent. By Definition 1.5, it is sufficient to bound the the spectral radius of the influence matrix $\Psi_\pi^{\sigma_\Lambda}$ for any feasible partial configuration $\sigma_\Lambda \in \Omega(\pi_\Lambda)$ specified on any subset $\Lambda \subset V$.

Let $D \in \mathbb{R}^{V \times V}$ be the diagonal matrix with $D(v, v) = \Delta_v$ for all V. By similarity, $D^{-1}\Psi_\pi^{\sigma_\Lambda}D$ has the same eigenvalues as $\Psi_\pi^{\sigma_\Lambda}$, and

$$\rho(\Psi_\pi^{\sigma_\Lambda}) = \rho(D^{-1}\Psi_\pi^{\sigma_\Lambda}D) \leq \left\|D^{-1}\Psi_\pi^{\sigma_\Lambda}D\right\|_\infty.$$

The next lemma follows immediately.

Lemma B.1. *Let $\Lambda \subset V$ and $\sigma_\Lambda \in \Omega(\pi_\Lambda)$. Let $\eta > 0$. If for every $r \in V \setminus \Lambda$,*

$$(99) \qquad \sum_{v \in V \setminus \Lambda} \Delta_v \cdot \Psi_\pi^{\sigma_\Lambda}(r, v) \leq \eta \cdot \Delta_r,$$

then it holds that $\rho(\Psi_\pi^{\sigma_\Lambda}) \leq \eta$.

Fix $\Lambda \subset V$ and $\sigma_\Lambda \in \Omega(\mu_\Lambda)$. To bound $\rho(\Psi_\pi^{\sigma_\Lambda})$, we only need to bound the weighted row sum of $\Psi_\pi^{\sigma_\Lambda}$ given by (99) (which corresponds to weighted total influence in π^{σ_Λ}). This follows from a two-step argument:

- First, show that there is a two-spin system $\mathcal{I}_T$ defined on a SAW-tree rooted at r that preserves the influence from r to other vertices in $\mathcal{I}$ (Appendix B.2.1).
- Next, show that for any two-spin system $\mathcal{I}_T$ on a a rooted tree, the α-contraction and the c-boundedness of ϕ implies that the weighted total influence from the root to all other vertices in $\mathcal{I}_T$ is always bounded by $\frac{2c}{\alpha} \cdot \Delta_r$ (Appendix B.2.2).

B.2.1. *Self-avoiding walk tree.* Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system on a connected graph $G = (V, E)$ with Gibbs distribution π. Assume a total ordering among all vertices in G. Given a vertex $r \in V$, let $T = T_{\text{SAW}}(G, r) = (V_{\text{SAW}}, E_{\text{SAW}})$ be the *self-avoiding walk (SAW) tree* rooted at r. The unique path between r and each vertex u in T is either a self-avoiding walk that ends at u or a self-avoiding walk with u being a cycle-closing vertex. The vertices in T enumerate all such walks in G and u is v's parent in T iff v's walk extends u's. The two-spin system $\mathcal{I}_T$ on T is defined as follow:

(1) A boundary condition is imposed to $\mathcal{I}_T$ on every cycle-closing leaf, such that a $+1$ is imposed on such a leaf if the cycle is formed from a smaller vertex to a larger vertex in the total ordering (and a -1 is imposed if otherwise). Denote this cycle-closing boundary condition on $\mathcal{I}_T$ as τ_{SAW}.
(2) $\mathcal{I}_T$ has the same edge interactions β, γ as $\mathcal{I}$, and for every free vertex v in T, the local field of $\hat{v}$ is set to be λ_v if $\hat{v}$ is a copy of v in T.

We denote the Gibbs distribution of $\mathcal{I}_T$ as π_{SAW}. For each $v \in V$, we denote by C_v the set of all free (not fixed by τ) copies of v in $\mathcal{I}_T$. For any partial configuration $\sigma_\Lambda \in \Omega(\pi_\Lambda)$ specified on a subset $\Lambda \subset V$, a corresponding partial configuration σ_{SAW} can be constructed in $\mathcal{I}_T$ by assigning σ_v to each copy $\hat{v}$ in C_v for any $v \in V$. Together, $\sigma_{\text{SAW}} \uplus \tau_{\text{SAW}}$ forms the boundary condition on Λ_{SAW} that is imposed on $\mathcal{I}_T$. With slightly abuse of the notation, we also use σ_Λ to denote this boundary condition $\sigma_{\text{SAW}} \uplus \tau_{\text{SAW}}$ on $\mathcal{I}_T$. We use $\Psi_G^{\sigma_\Lambda}$ and $\Psi_T^{\sigma_\Lambda}$ to denote $\Psi_\pi^{\sigma_\Lambda}$ and $\Psi_{\pi_{\text{SAW}}}^{\sigma_{\text{SAW}} \uplus \tau_{\text{SAW}}}$ respectively, when $T = T_{\text{SAW}}(G, r)$ and $\mathcal{I}_T$ are generated according to the above construction.

The following signed influence is introduced in [?].

Definition B.2 (pairwise signed influence). Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system on $G = (V, E)$ with Gibbs distribution π. Let $\Lambda \subset V$ and $\sigma_\Lambda \in \Omega(\mu_\Lambda)$. The *pairwise signed influence* is defined as follows. For every $u, v \in V$ such that $u \neq v$, let

$$I_\pi^{\sigma_\Lambda}(u, v) \triangleq \begin{cases} \pi_v^{\sigma_\Lambda, u \leftarrow +1}(+1) - \pi_v^{\sigma_\Lambda, u \leftarrow -1}(+1) & \text{if } \Omega(\mu_u^{\sigma_\Lambda}) = \{-1, +1\}, \\ 0 & \text{otherwise,} \end{cases}$$

and for every $u \in V$, let $I_\pi^{\sigma_\Lambda}(u, u) = 0$.

53

Here, for convenience, we also use $I_G^{\sigma_\Lambda}$ and $I_T^{\sigma_\Lambda}$ to denote $I_\pi^{\sigma_\Lambda}$ and $I_{\pi_{\mathrm{SAW}}}^{\sigma_\Lambda}$, respectively.

The following lemma is taken from [?], which was proved for two-spin systems with local fields.

Lemma B.3 ([? , Lemma 8]). *Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system on $G = (V, E)$ with Gibbs distribution π, where $0 \leq \beta \leq \gamma$, $\gamma > 0$, and $\lambda_v > 0$ for all $v \in V$. Let $r \in V$ be a vertex, and $\mathcal{I}_T$ the two-spin system obtained from $\mathcal{I}$ according to the SAW-tree transformation $T_{\mathrm{SAW}}(G, r)$. Let $\Lambda \subset V$ and $\sigma_\Lambda \in \Omega(\mu_\Lambda)$. For any $u \in V \setminus \Lambda$, it holds that*

$$I_G^{\sigma_\Lambda}(r, u) = \sum_{\hat{u} \in C_u} I_T^{\sigma_\Lambda}(r, \hat{u}).$$

Let $\Lambda \subset V$, $\sigma_\Lambda \in \Omega(\mu_\Lambda)$, and $r \in V \setminus \Lambda$. By the construction of $\mathcal{I}_T$ and Lemma B.3, the followings hold:

- for any $v \in V$, every free copy $\hat{v}$ in $\mathcal{I}_T$ has the same degree and local field as v in $\mathcal{I}$;
- it holds that

$$\sum_{v \in V \setminus \Lambda} \Delta_{G,v} \cdot \Psi_G^{\sigma_\Lambda}(r, v) = \sum_{v \in V \setminus \Lambda} \Delta_{G,v} \cdot \left| I_G^{\sigma_\Lambda}(r, v) \right| = \sum_{v \in V \setminus \Lambda} \Delta_{G,v} \cdot \left| \sum_{\hat{v} \in C_u} I_T^{\sigma_\Lambda}(r, \hat{v}) \right|$$

$$\leq \sum_{v \in V \setminus \Lambda} \sum_{\hat{v} \in C_u} \Delta_{G,v} \cdot \left| I_T^{\sigma_\Lambda}(r, \hat{v}) \right| = \sum_{\hat{v} \in V_{\mathrm{SAW}} \setminus \Lambda_{\mathrm{SAW}}} \Delta_{T,\hat{v}} \cdot \Psi_T^{\sigma_\Lambda}(r, \hat{v}),$$

where $\Delta_{G,v}$, $\Delta_{T,\hat{v}}$ denote the degree of v in G and $\hat{v}$ in tree $T = T_{\mathrm{SAW}}(G, r)$ respectively, and $V_{\mathrm{SAW}} \setminus \Lambda_{\mathrm{SAW}}$ is the set all free vertices in $\mathcal{I}_T$.

In order to get an upper bound for the weighted row sum of $\Psi_G^{\sigma_\Lambda}$ with respect to r, it is then sufficient for us to get an upper bound for the weighted row sum of $\Psi_T^{\sigma_\Lambda}$ with respect to r.

B.2.2. *Influence on trees.* We now bound the total influence on trees by assuming (α, c)-potential function. To work with SAW trees with boundary conditions, we slightly generalize the definition of (α, c)-potential function in Definition 8.7. Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system with local fields. Let $\phi : [-\infty, +\infty] \to [0, +\infty)$ be a function such that $\phi(y) > 0$ for any $y \in [-\infty, +\infty]$ that $|h(y)| > 0$, where the function h is defined in (77). Let $S \subseteq V$ be a subset. For any $\alpha \in (0, 1)$ and $c > 0$, we say ϕ is an (α, c)-*potential function with respect to $\mathcal{I}$ on S* if it satisfies the following conditions:

(1) (α-Contraction) For every $v \in S$ with $d_v \geq 1$ and every $(y_1, \ldots, y_{d_v}) \in [-\infty, +\infty]^{d_v}$, we have

$$\phi(y) \sum_{i=1}^{d_v} h^\phi(y_i) \leq 1 - \alpha,$$

where $y = H_{\lambda_v, d_v}(y_1, \ldots, y_d)$.

(2) (c-Boundedness) For every $u, v \in S$, every $y_u \in J_{\lambda_u, d_u}$ and $y_v \in J_{\lambda_v, d_v}$, we have

$$\phi(y_v) \cdot h^\phi(y_u) \leq \frac{2c}{\Delta_u + \Delta_v}.$$

Recall that $h^\phi : [-\infty, +\infty] \to [0, +\infty)$ is defined as that for any $y \in [-\infty, +\infty]$, $h^\phi(y) = 0$ if $h(y) = 0$, and if $h(y) \neq 0$,

$$h^\phi(y) = \frac{|h(y)|}{\phi(y)}.$$

The only difference from Definition 8.7 is that in the above definition, the α-Contraction and c-Boundedness properties are required to hold on a subset S. We will prove the following lemma.

Lemma B.4. *Let $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ be a two-spin system on a tree $T = (V, E)$ with Gibbs distribution π, where $0 \leq \beta \leq \gamma$, $\gamma > 0$, and $\lambda_v > 0$ for all $v \in V$. Let $r \in V$, $\Lambda \subseteq V \setminus \{r\}$ and $\sigma_\Lambda \in \Omega(\pi_\Lambda)$. For any*

$\alpha \in (0, 1)$ *and* $c > 0$, *if there is an* (α, c)-*potential function* ϕ *with respect to* $\mathcal{I}$ *on* $V \setminus \Lambda$, *then it holds that*

$$\sum_{v \in V \setminus \Lambda} \Delta_v \cdot \Psi_\pi^{\sigma_\Lambda}(r, v) \le \Delta_r \cdot \frac{2c}{\alpha}.$$

Fix $\Lambda \subseteq V \setminus \{r\}$, and $\sigma_\Lambda \in \Omega(\mu_\Lambda)$. Without loss of generality, we assume that Λ only contains leaves of T, since for any $v \in \Lambda$, due to conditional independence the descendants of such v can be removed without affecting the influence from the root. We use T_v to denote the subtree of T rooted at vertex $v \in V$, and $L_v(k) \subseteq V \setminus \Lambda$ for the set of all *free* vertices at distance k away from v in the subtree T_v. Moreover, we define

$$\Upsilon_\pi^{\sigma_\Lambda} \triangleq \Psi_\pi^{\sigma_\Lambda} + I,$$

where I denotes the identity matrix. We will prove the following lemma.

Lemma B.5. *Let* $c > 0$ *and* $\alpha \in (0, 1)$ *be two real numbers. If there is an* (α, c)-*potential function* ϕ *with respect to* $\mathcal{I}$ *on* $V \setminus \Lambda$, *then for any integer* $k \ge 1$, *it holds that*

$$\sum_{u \in L_r(k)} \Delta_u \Psi_\pi^{\sigma_\Lambda}(r, u) \le \sum_{u \in L_r(k)} \Delta_u \Upsilon_\pi^{\sigma_\Lambda}(r, u) \le 2c(1 - \alpha)^{k-1} \Delta_r.$$

Lemma B.4 is an easy consequence of Lemma B.5.

Proof of Lemma B.4. Note that $\Psi_T^{\sigma_\Lambda}(r, r) = 0$. It holds that

$$\sum_{u \in V_T} \Delta_u \cdot \Psi_T^{\sigma_\Lambda}(r, u) = \sum_{k=1}^{+\infty} \sum_{u \in L_r(k)} \Delta_u \cdot \Psi_T^{\sigma_\Lambda}(r, u) \le \Delta_r \cdot \sum_{k=1}^{\infty} 2c(1 - \alpha)^{k-1} = \frac{2c}{\alpha} \Delta_r,$$

where the inequality is due to Lemma B.5. $\qquad\square$

Let $u \in V$ be a vertex in the rooted tree T. Let $\mathcal{I}_u = \mathcal{T}_{T_u}$ denote the two-spin system induced by $\mathcal{I}_T$ on the subtree T_u rooted by u. Let ν be the Gibbs distribution associated with $\mathcal{I}_u$. Let τ denote the configuration obtained by restricting σ_Λ on T_u, formally, $\tau = \sigma_\Lambda(V_{T_u})$, where V_{T_u} denotes all vertices in T_u. We define the *marginal ratio* at u by

$$R_u = \frac{\nu_u^\tau(+1)}{\nu_u^\tau(-1)}.$$

It is well know [?] that each R_u can be calculated by the following tree recursion. For a vertex $u \in V$ whose local field is λ_u, denote the children of u as $w_1, w_2, \cdots, w_{d_u}$, where d_u is the number of u's children in T_u. The recursion of the marginal ratio is defined explicitly as follow:

$$R_u = \begin{cases} 0 & u \in \Lambda, \sigma_u = -1; \\ +\infty & u \in \Lambda, \sigma_u = +1; \\ F_{\lambda_u, d_u}(R_{w_1}, R_{w_2}, \cdots, R_{w_{d_u}}) & u \notin \Lambda, \end{cases}$$

where we recall that the tree recursion F_{λ_u, d_u} is defined as:

$$F_{\lambda_u, d_u}\left(R_{w_1}, R_{w_2}, \cdots, R_{w_{d_u}}\right) = \lambda_u \prod_{i=1}^{d_u} \frac{\beta R_{w_i} + 1}{R_{w_i} + \gamma}.$$

Now, we prove Lemma B.5. Recall that for two vertices $u, v \in V \setminus \Lambda$, if $u \ne v$, then

$$\Upsilon_\pi^{\sigma_\Lambda}(u, v) = \Psi_\pi^{\sigma_\Lambda}(u, v) = \max_{i, j \in \Omega(\pi_u^{\sigma_\Lambda})} d_{\text{TV}}\left(\pi_v^{\sigma_\Lambda, u \leftarrow i}, \pi_v^{\sigma_\Lambda, u \leftarrow j}\right)$$

gives the influence from u to v; and $\Upsilon_\pi^{\sigma_\Lambda}(w, w) = 1$ for any $w \in V \setminus \Lambda$.

The following lemma is taken from [?], which can be straightforwardly extended to two-spin systems with local fields.

Lemma B.6 ([? , Lemma B.2]). *Let $u, v, w \in V \setminus \Lambda$ be distinct vertices in tree T such that u is on the unique path from v to w. The signed influence matrix $I_\pi^{\sigma_\Lambda}$ defined in Definition B.2 satisfies*

$$I_\pi^{\sigma_\Lambda}(v, w) = I_\pi^{\sigma_\Lambda}(v, u) \cdot I_\pi^{\sigma_\Lambda}(u, w).$$

Note that for any $u, v \in V \setminus \Lambda$ with $u \neq v$, it holds that $\Upsilon_\pi^{\sigma_\Lambda} = \left| I_\pi^{\sigma_\Lambda}(u, v) \right|$, and for any $w \in V \setminus \Lambda$, it holds that $\Upsilon_\pi^{\sigma_\Lambda}(w, w) = 1$.

Corollary B.7. *Let $u, v, w \in V \setminus \Lambda$ be (not necessarily distinct) vertices in tree T such that u is on the unique path from v to w. It holds that*

$$\Upsilon_\pi^{\sigma_\Lambda}(v, w) = \Upsilon_\pi^{\sigma_\Lambda}(v, u) \cdot \Upsilon_\pi^{\sigma_\Lambda}(u, w).$$

The following lemma [?] was proved for two-spin systems with local fields.

Lemma B.8 ([? , Lemma 16]). *Let $v \in V \setminus \Lambda$ and $u \in V \setminus \Lambda$ a child of v in the subtree T_v. If $|\Omega(\pi_v^{\sigma_\Lambda})| = 1$, then $\Upsilon_\pi^{\sigma_\Lambda}(v, u) = 0$; otherwise, it holds that*

$$\Upsilon_\pi^{\sigma_\Lambda}(v, u) = |I_\pi^{\sigma_\Lambda}(v, u)| = |h(\log R_u)|,$$

where $h(y) = -\frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}$.

Now, we are ready to prove Lemma B.5.

Proof of Lemma B.5. Suppose there is a (α, c)-potential function ϕ with respect to $\mathcal{I}$ on $V \setminus \Lambda$. We denote the children of r by $u_1, u_2, \cdots, u_{\Delta_r}$. Then by Corollary B.7 and Lemma B.8, for any integer $k \geq 1$,

$$\sum_{v \in L_r(k)} \Delta_v \cdot \Upsilon_\pi^{\sigma_\Lambda}(r, v) = \sum_{i=1}^{\Delta_r} \Upsilon_\pi^{\sigma_\Lambda}(r, u_i) \sum_{v \in L_{u_i}(k-1)} \Upsilon_\pi^{\sigma_\Lambda}(u_i, v) \Delta_v$$

$$\text{(by Lemma B.8)} \quad \leq \sum_{i=1}^{\Delta_r} \left| h(\log R_{u_i}) \right| \sum_{v \in L_{u_i}(k-1)} \Upsilon_\pi^{\sigma_\Lambda}(u_i, v) \Delta_v$$

$$(100) \qquad\qquad \leq \sum_{i=1}^{\Delta_r} h^\phi(\log R_{u_i}) \sum_{v \in L_{u_i}(k-1)} \phi(\log R_{u_i}) \Upsilon_\pi^{\sigma_\Lambda}(u_i, v) \Delta_v,$$

where (100) holds because of the fact that $\phi(y) > 0$ for all $y \in [-\infty, +\infty]$ that $|h(y)| > 0$ and the definition of $h^\phi(y)$ such that $h^\phi(y) = 0$ if $h(y) = 0$ and $h^\phi(y) = \frac{|h(y)|}{\phi(y)}$ if $h(y) \neq 0$. We have for any integer $k \geq 1$,

$$(101) \qquad \sum_{v \in L_r(k)} \Delta_v \cdot \Upsilon_\pi^{\sigma_\Lambda}(r, v) \leq \Delta_r \cdot \max_{1 \leq i \leq \Delta_r} \left\{ h^\phi(\log R_{u_i}) \right\} \cdot \max_{1 \leq i \leq \Delta_r} \left\{ \sum_{v \in L_{u_i}(k-1)} \phi(\log R_{u_i}) \Upsilon_\pi^{\sigma_\Lambda}(u_i, v) \Delta_v \right\},$$

To bound the right-hand-side of (101), we claim that for all $w \in V \setminus (\{r\} \cup \Lambda)$ and $k \geq 0$,

$$(102) \qquad \sum_{v \in L_w(k)} \Delta_v \cdot \phi(\log R_w) \Upsilon_\pi^{\sigma_\Lambda}(w, v) \leq \max_{v \in L_w(k)} \{ \Delta_v \cdot \phi(\log R_v) \} \cdot (1 - \alpha)^k .$$

Combining (101) and (102), we have

$$\sum_{v \in L_r(k)} \Delta_v \cdot \Upsilon_\pi^{\sigma_\Lambda}(r, v) \leq \Delta_r \cdot \max_{1 \leq i \leq \Delta_r} \left\{ h^\phi(\log R_{u_i}) \right\} \cdot \max_{1 \leq i \leq \Delta_r} \max_{v \in L_{u_i}(k-1)} \{ \Delta_v \cdot \phi(\log R_v) \} \cdot (1 - \alpha)^{k-1}$$

$$= \Delta_r \cdot \max_{1 \leq i \leq \Delta_r} \left\{ h^\phi(\log R_{u_i}) \right\} \cdot \max_{v \in L_r(k)} \{ \Delta_v \cdot \phi(\log R_v) \} \cdot (1 - \alpha)^{k-1}$$

$$= \Delta_r \cdot \max_{u \in L_r(1), v \in L_r(k)} \left\{ \Delta_v \cdot h^\phi(\log R_{u_i}) \cdot \phi(\log R_v) \right\} \cdot (1 - \alpha)^{k-1}$$

$$(\star) \quad \leq \Delta_r \cdot 2c \cdot (1 - \alpha)^{k-1}.$$

Inequality ($\star$) holds because of the boundedness property and the fact that $u, v \in V \setminus \Lambda$ are free variables.

We now use induction on k to prove (102). The base case is $k = 0$. We have

$$\Delta_w \cdot \phi(\log R_w)\Upsilon_\pi^{\sigma_\Lambda}(w, w) = \Delta_w \cdot \phi(\log R_w),$$

which holds because $\Upsilon_\pi^{\sigma_\Lambda}(w, w) = 1$.

Now, suppose (102) holds for some integer $k - 1 \geq 0$. Fix $w \in V \setminus (\Lambda \cup \{r\})$. Let $w_1, w_2, \ldots, w_{d_w}$ denote the children of w in T_w. Since $w \neq r$, we have $1 \leq d_w < \Delta$. We assume $d_w > 0$ because (102) holds trivially if $d_w = 0$. By Corollary B.7 and Lemma B.8, we have

$$\sum_{v \in L_w(k)} \Delta_v \cdot \phi(\log R_w)\Upsilon_\pi^{\sigma_\Lambda}(w, v) = \sum_{i=1}^{d_w} \phi(\log R_w)\Upsilon_\pi^{\sigma_\Lambda}(w, w_i) \sum_{v \in L_{w_i}(k-1)} \Upsilon_\pi^{\sigma_v}(w_i, v) \cdot \Delta_v$$

$$\leq \sum_{i=1}^{d_w} \phi(\log R_w) \left| h(\log R_{w_i}) \right| \sum_{v \in L_{w_i}(k-1)} \Upsilon_\pi^{\sigma_v}(w_i, v) \cdot \Delta_v.$$

By a similar argument in (100), we have

$$\sum_{v \in L_w(k)} \Delta_v \cdot \phi(\log R_w)\Upsilon_\pi^{\sigma_\Lambda}(w, v) \leq \sum_{i=1}^{d_w} \phi(\log R_w)h^\phi(\log R_{w_i}) \sum_{v \in L_{w_i}(k-1)} \phi(\log R_{w_i})\Upsilon_\pi^{\sigma_v}(w_i, v) \cdot \Delta_v,$$

By the induction hypothesis, it holds that

$$\sum_{v \in L_w(k)} \Delta_v \cdot \phi(\log R_w)\Upsilon_\pi^{\sigma_\Lambda}(w, v) \leq \sum_{i=1}^{d_w} \phi(\log R_w)h^\phi(\log R_{w_i}) \max_{v \in L_{w_i}(k-1)} \{\Delta_v \cdot \phi(\log R_v)\} \cdot (1 - \alpha)^{k-1}$$

$$\leq \max_{v \in L_w(k)} \{\Delta_v \cdot \phi(\log R_v)\} \cdot (1 - \alpha)^{k-1} \cdot \sum_{i=1}^{d_w} \phi(\log R_w)h^\phi(\log R_{w_i})$$

$$\leq \max_{v \in L_w(k)} \{\Delta_v \cdot \phi(\log R_v)\} \cdot (1 - \alpha)^k,$$

where the last inequality holds because $w \in V \setminus \Lambda$ and $d_w \geq 1$, and ϕ satisfies the α-Contraction property on $V \setminus \Lambda$. $\square$

Proof of Theorem 8.8. Recall that $\mathcal{I} = (V, E, \beta, \gamma, (\lambda_v)_{v \in V})$ is a two-spin system with Gibbs distribution π where $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\lambda_v > 0$ for all $v \in V$, and ϕ is an (α, c)-function with respect to $\mathcal{I}$. For any $r \in V$, $\Lambda \subseteq V \setminus \{r\}$, and $\sigma_\Lambda \in \Omega(\pi_\Lambda)$, let $\mathcal{I}_T$ denote the two-spin system on the SAW tree $T_{\text{SAW}}(G, r)$ whose Gibbs distribution is π_{SAW}. Recall that we use $\Psi_T^{\sigma_\Lambda}$ to denote $\Psi_{\pi_{\text{SAW}}}^{\sigma_\Lambda}$ and $\Psi_G^{\sigma_\Lambda}$ to denote $\Psi_\pi^{\sigma_\Lambda}$ respectively. Then, by Lemma B.3, we have

$$(103) \qquad \sum_{v \in V \setminus \Lambda} \Delta_{G,v} \cdot \Psi_G^{\sigma_\Lambda}(r, v) \leq \sum_{\hat{v} \in V_{\text{SAW}} \setminus \Lambda_{\text{SAW}}} \Delta_{T,\hat{v}} \cdot \Psi_T^{\sigma_\Lambda}(r, \hat{v}),$$

By the SAW-tree construction, for any $v \in V$, if $\hat{v}$ is a free copy of v in $\mathcal{I}_T$, then v and $\hat{v}$ have the same local field and degree. It is straightforward to verify that the function ϕ is also an (α, c)-function with respect to $\mathcal{I}_T$ on $V_{\text{SAW}} \setminus \Lambda_{\text{SAW}}$. Lemma B.4 can be applied to $\mathcal{I}_T$, and it holds that

$$(104) \qquad \sum_{\hat{v} \in V_{\text{SAW}} \setminus \Lambda_{\text{SAW}}} \Delta_{T,\hat{v}} \cdot \Psi_T^{\sigma_\Lambda}(r, \hat{v}) \leq \Delta_r \cdot \frac{2c}{\alpha}.$$

Finally, Theorem 8.8 is proved by combining (103) and (104) with Lemma B.1. $\square$

B.3. Proof of Lemma 8.11. Recall the definition of the tree recursion for marginal ratios $F_{\lambda,d}(x_1, \ldots, x_d)$ in (75). For $\vec{x} = (x_1, \ldots, x_d) \in [0, +\infty]^d$, let

$$\alpha_d(x_1, \ldots, x_d) \triangleq \sqrt{\frac{(1 - \beta\gamma)F_{\lambda,d}(\vec{x})}{(\beta F_{\lambda,d}(\vec{x}) + 1)(F_{\lambda,d}(\vec{x}) + \gamma)}} \sum_{i=1}^d \sqrt{\frac{(1 - \beta\gamma)x_i}{(\beta x_i + 1)(x_i + \gamma)}}.$$

Recall the definition of the tree recursion for log-marginal-ratios $H_{\lambda,d}(y_1, \ldots, y_d)$ in (76), function $h(y)$ in (77) and $h^\phi(y)$ (78). For any integer $d \geq 1$, any $y_1, y_2, \ldots, y_d \in [-\infty, +\infty]^d$ and $y = H_{\lambda,d}(y_1, y_2, \ldots, y_d)$, it holds that

$$\text{(105)} \qquad \phi(y) \sum_{i=1}^d h^\phi(y_i) = \sum_{i=1}^d \sqrt{|h(y)h(y_i)|} = \alpha_d(e^{y_1}, \ldots, e^{y_d}).$$

For $x \in [0, +\infty]$, let $\alpha_d(x)$ denote the symmetric version of $\alpha_d(x_1, \ldots, x_d)$, specifically,

$$\alpha_d(x) \triangleq \alpha_d(\underbrace{x, \ldots, x}_{d}) = \sqrt{f_d(F_d(x))}\sqrt{f_d(x)},$$

where $F_d(x) \triangleq \lambda\left(\frac{\beta x + 1}{x + \gamma}\right)^d$, defined in (8), is the symmetric version of the tree recursion for marginal ratios $F_{\lambda,d}(\vec{x})$; and $f_d(d) \triangleq \frac{d(1 - \beta\gamma)x}{(\beta x + 1)(x + \gamma)}$ is defined in (9). The following symmetrization was well known.

Lemma B.9 ([?, Lemma 13]). *Let $d \geq 1$ be an integer, and let β, γ, λ be real numbers satisfying that $0 \leq \beta \leq \gamma, \gamma > 0, \lambda \geq 0$ and $\beta\gamma < 1$. For any $x_1, x_2, \ldots, x_d \in [0, +\infty)$, there exists $\overline{x} \in [0, +\infty)$ such that*

$$\alpha_d(x_1, x_2, \ldots, x_d) \leq \alpha_d(\overline{x}).$$

Furthermore, it was known that $\alpha_d(x)$ is bounded by $\sqrt{f_d(\hat{x}_d)}$ when (β, γ, λ) is d-unique.

Lemma B.10 ([?]). *Let $d \geq 1$ be an integer, and let β, γ, λ be real numbers satisfying that $0 \leq \beta \leq \gamma, \gamma > 0, \lambda \geq 0$ and $\beta\gamma < 1$. If (β, γ, λ) is d-unique with gap δ, then for any $x \in [0, +\infty)$,*

$$\alpha_d(x) = \sqrt{f_d(F_d(x))}\sqrt{f_d(x)} \leq \sqrt{f_d(\hat{x}_d)} \leq \sqrt{1 - \delta},$$

where $\hat{x}_d$ denotes the unique positive fixed point of $F_d(x) = \lambda\left(\frac{\beta x + 1}{x + \gamma}\right)^d$ and $f_d(\hat{x}_d) = |F'_d(\hat{x}_d)| = \frac{d(1 - \beta\gamma)\hat{x}}{(\beta\hat{x} + 1)(\hat{x} + \gamma)}$.

Remark B.11. In [?], the lemma was proved by assuming the up-to-Δ uniqueness (but the conclusion was also stronger which held for all $1 \leq d < \Delta$). Lemma B.10 can be proved by going through that proof.

Proof. By taking derivative

$$\text{(106)} \qquad \alpha'_d(x) = c(x) \cdot \left(\frac{\gamma - \beta x^2}{d(1 - \beta\gamma)x} - \frac{\gamma - \beta F_d(x)^2}{(\beta F_d(x) + 1)(F_d(x) + \gamma)}\right),$$

where $c(x) > 0$ is a function that is always positive for $x \in [0, +\infty)$. Note that $\frac{\gamma - \beta x^2}{d(1 - \beta\gamma)x}$ is decreasing in x and ranges from $+\infty$ to $-\infty$ and $F_d(x)$ is decreasing in x and has a bounded range $[\lambda\beta^d, \lambda\gamma^{-d}]$. Then the equation

$$\text{(107)} \qquad \frac{\gamma - \beta x^2}{d(1 - \beta\gamma)x} = \frac{\gamma - \beta F_d(x)^2}{(\beta F_d(x) + 1)(F_d(x) + \gamma)}.$$

has a unique positive solution $x_d \in (0, +\infty)$, at which $\alpha_d(x)$ achieves its maximum. Substituting $(\beta F_d(x_d) + 1)(F_d(x_d) + \gamma) = \frac{d(1 - \beta\gamma)x_d(\gamma - \beta F_d(x_d)^2)}{\gamma - \beta x_d^2}$ in $\alpha_d(x_d)$, we have

$$\alpha_d(x_d) = \sqrt{d(1 - \beta\gamma) \cdot \frac{(\gamma - \beta x_d^2)}{(\beta x_d + 1)(x_d + \gamma)} \cdot \frac{F_d(x_d)}{(\gamma - \beta F_d(x_d)^2)}} \triangleq \tilde{\alpha}_d(x_d).$$

58

We then claim that if (β, γ, λ) is d-unique, then $\tilde{\alpha}_d(x_d) \leq \tilde{\alpha}_d(\hat{x}_d)$ for the fixed point $\hat{x}_d = F_d(\hat{x}_d)$. To see that this is sufficient to prove the lemma, note that $\alpha_d(x) \leq \alpha_d(x_d) = \tilde{\alpha}_d(x_d) \leq \tilde{\alpha}_d(\hat{x}_d)$, and by substituting $F_d(\hat{x}_d) = \hat{x}_d$, we have

$$\tilde{\alpha}_d(\hat{x}_d) = \sqrt{\frac{d(1 - \beta\gamma)\hat{x}_d}{(\beta\hat{x}_d + 1)(\hat{x}_d + \gamma)}} = \sqrt{f_d(\hat{x}_d)}.$$

Next, we prove this claim. It is sufficient to show that $\tilde{\alpha}_d(x)$ is decreasing in $x \in [\hat{x}_d, x_d]$ if $\hat{x}_d \leq x_d$ and $\tilde{\alpha}_d(x)$ is increasing in $x \in [x_d, \hat{x}_d]$ if $\hat{x}_d > x_d$.

- Case 1: $\hat{x}_d \leq x_d$. In this case, according to (106), $\alpha'(\hat{x}_d) \geq 0$. Note that

$$\alpha'(\hat{x}_d) = c(\hat{x}_d)(\gamma - \beta\hat{x}_d^2)\left(\frac{1}{d(1 - \beta\gamma)\hat{x}_d} - \frac{1}{(\beta\hat{x}_d + 1)(\hat{x}_d + \gamma)}\right),$$

  where $c(\hat{x}_d) > 0$. Due to the d-uniqueness of (β, γ, λ), we have $|f_d'(\hat{x}_d)| = \frac{d(1-\beta\gamma)\hat{x}_d}{(\beta\hat{x}_d+1)(\hat{x}_d+\gamma)} < 1$, that is, $\frac{1}{d(1-\beta\gamma)\hat{x}_d} - \frac{1}{(\beta\hat{x}_d+1)(\hat{x}_d+\gamma)} > 0$. Therefore $\alpha'(\hat{x}_d) \geq 0$ means that $\gamma - \beta\hat{x}_d^2 \geq 0$. Since $F_d(x)$ is monotonically decreasing in x and $\hat{x}_d$ is its fixed point, we have

$$\gamma - \beta F_d(x_d)^2 \geq \gamma - \beta F_d(\hat{x}_d)^2 = \gamma - \beta\hat{x}_d^2 \geq 0.$$

  Since x_d satisfies (107), $\gamma - \beta x_d^2$ and $\gamma - \beta F_d(x_d)^2$ must be simultaneously positive or negative, thus it also holds that $\gamma - \beta x_d^2 \geq 0$. Then both $\frac{(\gamma - \beta x^2)}{(\beta x + 1)(x + \gamma)}$ and $\frac{F_d(x)}{(\gamma - \beta F_d(x)^2)}$ are positive and monotonically decreasing in $x \in [\hat{x}_d, x_d]$. Therefore, $\tilde{\alpha}_d(x_d) \leq \tilde{\alpha}_d(\hat{x}_d)$.
- Case 2: $\hat{x}_d > x_d$. By symmetry, we have $\gamma - \beta F_d(\hat{x}_d)^2 = \gamma - \beta\hat{x}_d^2 < 0$, $\gamma - \beta F_d(x_d)^2 < 0$, and $\gamma - \beta x_d^2 < 0$. Thus both $\frac{(\gamma - \beta x^2)}{(\beta x + 1)(x + \gamma)}$ and $\frac{F_d(x)}{(\gamma - \beta F_d(x)^2)}$ are negative and monotonically decreasing in $x \in [x_d, \hat{x}_d]$, hence their product is positive and increasing in $x \in [x_d, \hat{x}_d]$. Therefore, $\tilde{\alpha}_d(x_d) \leq \tilde{\alpha}_d(\hat{x}_d)$.

$\square$

Proof of Lemma 8.11. We now prove Lemma 8.11. Given $y_1, y_2, \ldots, y_d \in [-\infty, +\infty]$, define

$$x_i = e^{y_i}.$$

If all $y_i \in [-\infty, +\infty)$, then all $x_i \in \mathbb{R}_{\geq 0}$. By (105), Lemma B.9 and Lemma B.10, it holds that

$$(108) \qquad \phi(y) \sum_{i=1}^{d} h^{\phi}(y_i) = \alpha_d(e^{y_1}, \ldots, e^{y_i}) = \alpha_d(x_1, \ldots, x_d) \leq \sqrt{1 - \delta}.$$

The lemma follows.

Suppose there is a subset $S \subseteq [d]$ such that $y_i = +\infty$ for all $i \in S$ and $y_j \in [-\infty, +\infty)$ for all $j \in [d] \setminus S$. In this case, $x_i = +\infty$ for all $i \in S$, and $x_j \in \mathbb{R}_{\geq 0}$ for all $j \in [d] \setminus S$. Define a function

$$g(x) = \alpha_d(z_1, z_2, \ldots, z_d),$$

where $z_i = x$ for all $i \in S$, and $z_j = x_j = e^{y_j} \in [0, +\infty)$ for all $j \in [d] \setminus S$. By (105) and definition of $\alpha_d(x_1, \ldots, x_d)$, it is straightforward to verify $\lim_{x \to \infty} g(x)$ exists and it holds that

$$\alpha_d(e^{y_1}, \ldots, e^{y_d}) = \alpha_d(x_1, \ldots, x_d) = \lim_{x \to \infty} g(x).$$

By (108), for any $x \in [0, +\infty)$, we have $g(x) \leq \sqrt{1 - \delta}$. Hence

$$\phi(y) \sum_{i=1}^{d} h^{\phi}(y_i) = \lim_{x \to \infty} g(x) \leq \sqrt{1 - \delta}. \qquad \square$$